PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/31, 15/62, C07K 14/295, 16/12, 19/00, A01K 67/027, A61K 39/118, G01N 33/53, C12Q 1/68 — A2

(11) International Publication Number: WO 99/27105

(43) International Publication Date: 3 June 1999 (03.06.99)

(21) International Application Number: PCT/IB98/01890

(22) International Filing Date: 20 November 1998 (20.11.98)

(30) Priority Data:
97/14673 21 November 1997 (21.11.97) FR
60/107,078 4 November 1998 (04.11.98) US

(71) Applicant (for all designated States except US): GENSET [FR/FR]; 24, rue Royale, F-75008 Paris (FR).

(72) Inventor; and
(75) Inventor/Applicant (for US only): GRIFFAIS, Rémy [FR/FR]; 51, boulevard Romain Roland, F-92120 Montrouge (FR).

(74) Agents: MARTIN, Jean-Jacques et al.; Cabinet Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: *CHLAMYDIA PNEUMONIAE* GENOMIC SEQUENCE AND POLYPEPTIDES, FRAGMENTS THEREOF AND USES THEREOF, IN PARTICULAR FOR THE DIAGNOSIS, PREVENTION AND TREATMENT OF INFECTION

(57) Abstract

The subject of the invention is the genomic sequence and the nucleotide sequences encoding polypeptides of *Chlamydia pneumoniae*, such as cellular envelope polypeptides, which are secreted or specific, or which are involved in metabolism, in the replication process or in virulence, polypeptides encoded by such sequences, as well as vectors including the said sequences and cells or animals transformed with these vectors. The invention also relates to transcriptional gene products of the *Chlamydia pneumoniae* genome, such as, for example, antisense and ribozyme molecules, which can be used to control growth of the microorganism. The invention also relates to methods of detecting these nucleic acids or polypeptides and kits for diagnosing *Chlamydia pneumoniae* infection. The invention also relates to a method of selecting compounds capable of modulating bacterial infection and a method for the biosynthesis or biodegradation of molecules of interest using the said nucleotide sequences or the said polypeptides. The invention finally comprises, pharmaceutical, in particular vaccine, compositions for the prevention and/or treatment of bacterial, in particular *Chlamydia pneumoniae*, infections.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

CHLAMYDIA PNEUMONIAE GENOMIC SEQUENCE AND POLYPEPTIDES, FRAGMENTS THEREOF AND USES THEREOF, IN PARTICULAR FOR THE DIAGNOSIS, PREVENTION AND TREATMENT OF INFECTION

The subject of the invention is the genomic sequence and the nucleotide sequences encoding polypeptides of *Chlamydia pneumoniae*, such as cellular envelope polypeptides, which are secreted or specific, or which are involved in metabolism, in the replication process or in virulence, polypeptides encoded by such sequences, as well as vectors including the said sequences and cells or animals transformed with these vectors. The invention also relates to transcriptional gene products of the *Chlamydia pneumoniae* genome, such as, for example, antisense and ribozyme molecules, which can be used to control growth of the microorganism. The invention also relates to methods of detecting these nucleic acids or polypeptides and kits for diagnosing *Chlamydia pneumoniae* infection. The invention also relates to a method of selecting compounds capable of modulating bacterial infection and a method for the biosynthesis or biodegradation of molecules of interest using the said nucleotide sequences or the said polypeptides. The invention finally comprises, pharmaceutical, in particular vaccine, compositions for the prevention and/or treatment of bacterial, in particular *Chlamydia pneumoniae*, infections.

Comparative analysis of the sequence of the gene encoding the ribosomal 16S RNA has been widely used for the phylogenetic study of prokaryotes. This approach has made it possible to classify the Chlamydiae among the eubacteria, among which they represent a well-isolated group, with, nevertheless, a very weak link with the planctomyces. The Chlamydiae thus exhibit some unique characteristics within the eubacteria, in particular their development cycle and the structure of their membranes. They have a unique two-phase cell cycle: the elementary body, a small extracellular form, attaches to the host and is phagocytosed; in the phagosome, it is converted to the replicative intracellular form, the reticulate body. The Chlamydiae are obligate intracellular bacteria which multiply in eukaryotic cells at the expense of their energy reserves and nucleotide pools; they are responsible for a wide variety of diseases in mammals and birds. The Chlamydiae are the only members of the order Chlamydiales, of the family Chlamydiaceae and of the genus Chlamydia. Within the genus *Chlamydia*, four species are currently described: *Chlamydia trachomatis*, *Chlamydia psittaci*, *Chlamydia pneumoniae* and *Chlamydia pecorum*. These bacteria are grouped together and share biological and biochemical properties. Among them, only the first three infect humans, *Chlamydia pecorum* being a pathogen of ruminants.

The species *Chlamydia psittaci* infects many animals, in particular birds, and is transmissible to humans. It is responsible for atypical pneumonia, for hepatic and renal dysfunction, for endocarditis and for conjunctivitis.

The species *Chlamydia trachomatis* is the best characterized. Besides a murine strain, it is divided into two groups which are distinguishable by the nature of the diseases for which they are responsible: trachoma, genital attack and venereal lymphogranulomatosis. There are fifteen human serotypes of *Chlamydia trachomatis* (A, K) and LGV (L1, L2, L3). Strains A to C are mainly found in eye infections, whereas strains D to K and LGV are essentially responsible for genital entry infections. It should be mentioned that the LGV strains are responsible for systemic diseases. Historically, it was in 1906 that Halberstaeder and Von Provaseck discovered, in trachoma patients, the presence of inclusions in the cytoplasm of the cells derived from conjunctival scrapings. In 1940, Rake and Jones described these same inclusions in certain cells obtained by puncturing the ganglia from a patient suffering from venereal granulomatosis. Characterization of the *Chlamydia trachomatis* microorganism was only successfully carried out in 1957, after a series of isolations in cell cultures.

It was in 1983 that *Chlamydia pneumoniae* was recognized as a human pathogen (Grayston JT et al., 1986); since then, special attention has been paid to this bacterium and it is estimated (Gaydos CA et al., 1994) that 10% of pneumonias, and 5% of bronchitides and sinusites are attributable to *Chlamydia pneumoniae* (Aldous MB et al., 1992). More recently, the association of this bacterium with the pathogenesis of asthmatic disease and of cardiovascular impairments is increasingly of interest.

Serological studies have made it possible to observe that *Chlamydia pneumoniae* infection is common in children between 5 and 16 years of age. Before this age, it is rare to find antibodies; the increase in the number of individuals carrying antibodies is then correlated with age up to 20 years. Accordingly, 50% of adults are carriers of antibodies, it being possible for this prevalence to be as high as 75%. These figures are all the more striking since a first infection induces antibody levels of which the persistence over time is limited to 3 or at most 5 years, which suggests frequent reinfection during the entire lifespan. The annual seroconversion rate is about 8% between 8 and 12 years and about 6% between 12 and 16 years (Haidl et al., 1994). Before the age of 15 years, the seroprevalence of the disease is identical between both sexes. After this age, men are more frequently infected than women; this is true in all regions worldwide where such studies have been carried out.

These infections are geographically highly widespread, as shown by numerous studies carried out throughout the world (Kanamoto Y et al., 1991; Tong CY et al., 1993). Developed countries of the north such as Canada, Denmark and Norway have the lowest infection rates; conversely, the highest prevalence rates are found in the less developed countries of tropical regions where the infection may occur before the age of 5 years.

Humans are the only known reservoir for *Chlamydia pneumoniae* and it is probable that the infection is caused by direct transmission, respiratory secretions probably being responsible for this low-yield transmission (Aldous et al., 1992). The chain of transmission may also appear to be indirect (Kleemola M et al., 1988), suggesting that the infection is caused by an effective transmission, but also that asymptomatic carriers exist, which could explain the high prevalence of the disease.

Other studies (Mordhorst CH et al., 1992) show that the efficiency of the transmission varies according to the individuals and list cases of infection affecting all or the majority of members of one family or of a group of families. The period of incubation is several weeks, significantly longer in this regard than that of many other respiratory pathogenic agents. Although under conditions of high relative humidity the infectivity of *Chlamydia pneumoniae* in the open air decreases rapidly, suggesting a direct mode of transmission under these conditions, it is probable that the transmission occurs in some cases indirectly since the microorganism can survive for up to 30 hours in a hostile environment (Falsey et al., 1993).

Clinical manifestations due to *Chlamydia pneumoniae* are essentially respiratory diseases. Pneumonia and bronchitis are the most frequent because they are clinically patent: since etiological diagnosis is evoked in this case, the infectious agent is identified. The asymptomatic diseases are probably numerous (Grayston JT et al., 1992; Grayston JT et al., 1986; Thom DH et al., 1990). The disease then progresses via bronchitis or pneumonia; fever is absent at the time of examination but is sometimes reported by the patient. The degree of seriousness of the disease is variable and in hospitalized patients, it is common to observe pleural effusion; a generalized infection may also be observed and, in severe cases, anatomicopathological examination shows *Chlamydia pneumoniae* diseases.

Other syndromes such as sinusitis (Hashiguchi K et al., 1992), purulent otitis media (Ogawa H et al., 1992), or pharyngitis (Huovinen P et al., 1989) have been described, as well as infections with respiratory impairments similar to asthma (Hahn DL et al., 1991). *Chlamydia pneumoniae* has also been associated with sarcoidosis, with erythema nodosum (Sundelof et al., 1993) and one case of Guillain-Barré syndrome has even been described (Haidl et al., 1992). The involvement of *Chlamydia pneumoniae* in Reiter's syndrome has also been evaluated (Braun J et al., 1994).

The association of *Chlamydia pneumoniae* with coronary diseases and with myocardial infarction was first suspected from the observation of the high antibody level in 71% of patients having a heart disease (Shor A et al., 1992; Kuo CC et al., 1993; Puolakkainen M et al., 1993; Thomas GN et al., 1997). Studies carried out in several countries have shown similar results in patients with atheromatous impairments (Shor A et al., 1992; Kuo CC et al., 1993; Puolakkainen M et al., 1993; Grayston JT et al., 1996; Casas-Ciria J et al., 1996; Thomas GN et al., 1997; Jackson LA et al., 1997) and in patients with carotid impairments. Anatomicopathological and microbiological studies have detected *Chlamydia pneumoniae* in the vessels. The electron microscope has made it possible to visualize the bacterium (Ladany S et al., 1989), which has in fact been demonstrated by other techniques such as PCR (Campbell LA et al., 1992; Kuo CC et al., 1993; Kuo CC et al., 1988). It also appears that the bacterium is more frequently found in old atheromatous lesions. Other studies carried out on young subjects from 15 to 35 years have given the opportunity to study the coronary arteries of people without atherosclerosis, this observation not being possible in older subjects (the

onset of the atheromatous disease is early). In these young subjects, the PCR studies did not find *Chlamydia pneumoniae* in subjects free of atheromatous disease, but revealed the presence of *Chlamydia pneumoniae* in two of the eleven subjects who showed early lesions and in six of the seven subjects who developed atheroma plaques. These studies therefore show that the atheroma plaque is very strongly correlated with the presence of *Chlamydia pneumoniae*, but the role played by the bacterium in vascular pathology is not yet defined.

The data relating to controlled clinical studies analysing the effect of treatments in *Chlamydia pneumoniae* infections are limited in number. Unlike penicillin, ampicillin or the sulphamides, erythromycin, tetracycline or doxycycline show an antibiotic activity *in vitro* against *Chlamydia pneumoniae*. However, a treatment at high doses should be continued for several weeks in order to avoid a recurrence of the infection. Accordingly, the use of two new macrolides, clarithromycin and azithromycin, whose diffusion, bioavailability and half-life allow shorter and better tolerated cures, is nowadays preferred. In the absence of definitive proof based on the results of clinical studies, an effective, without recurrences, and well-tolerated treatment of *Chlamydia pneumoniae* infections therefore remains desirable.

An even more important need up until now relates to a specific and sensitive diagnosis, which can be carried out conveniently and rapidly, allowing early screening for the infection. Methods based on *Chlamydia pneumoniae* culture are slow and require a considerable know-how because of the difficulty involved in the collection, preservation and storage of the strain under appropriate conditions. Methods based on antigen detection (EIA, DFA) or on nucleic acid amplification (PCR) provide tests which are more suitable for laboratory practice. A reliable, sensitive and convenient test, which allows distinction between serogroups and a fortiori between *Chlamydia pneumoniae* species is therefore highly desirable.

This is all the more important since the symptoms of *Chlamydia pneumoniae* infection appear slowly, since all the pathologies associated with these infections have not yet been identified, and since, as has been mentioned above, an association is suspected between these infections and serious chronic infections, asthma or atherosclerosis.

No vaccine is yet available against *Chlamydia pneumoniae*: this is due to the labile nature of the antigens specific to the strain, which has so far prevented their specific identification.

Although the number of studies and of animal models developed is high, the antigens used have not induced sufficient protective immunity to lead to the development of human vaccines. In the case of *Chlamydia pneumoniae*, the role of the immune defense in the physiology and pathology of the disease should probably be understood in order to develop satisfactory vaccines.

More detailed information relating to the biology of these strains, their interactions with their hosts, the associated phenomena of infectivity and those of escaping the immune defenses of the host in particular, and finally their involvement in the development of the these associated pathologies, will allow a better understanding of these mechanisms. In the light of the preceding text which shows

in particular the limitations of the means of controlling *Chlamydia pneumoniae* infection, it is therefore at present essential, on the one hand, to develop molecular tools, in particular from a better genetic knowledge of *Chlamydia pneumoniae*, but also to develop new preventive and therapeutic treatments, new diagnostic methods and new vaccine strategies which are specific, effective and tolerated. This is precisely the object of the present invention.

The subject of the present invention is the nucleotide sequence having the sequence SEQ ID No. 1 of the *Chlamydia pneumoniae* genome. However, the invention is not limited to SEQ ID No. 1, but encompasses genomes and nucleotides encoding polypeptides of strain variants, polymorphisms, allelic variants, and mutants.

Thus, the subject of the present invention encompasses nucleotide sequences characterized in that they are chosen from:

a) the nucleotide sequence of SEQ ID No. 1, a nucleotide sequence exhibiting at least 99.9% identity with the sequence SEQ ID No. 1, the nucleotide sequence of the genomic DNA contained within ATCC Deposit No. ___, the nucleotide sequence of a clone insert within ATCC Deposit No. __;

b) a nucleotide sequence homologous to the sequence SEQ ID No. 1;

c) a polynucleotide sequence that hybridizes to the nucleotide sequence of a) under conditions of high or intermediate stringency as described below:

(i) By way of example and not limitation, procedures using conditions of high stringency are as follows: Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65EC in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 μg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65EC, the preferred hybridization temperature, in prehybridization mixture containing 100 μg/ml denatured salmon sperm DNA and 5-20 X 10^6 cpm of ^{32}P-labeled probe. Alternatively, the hybridization step can be performed at 65EC in the presence of SSC buffer, 1 x SSC corresponding to 0.15M NaCl and 0.05 M Na citrate. Subsequently, filter washes can be done at 37EC for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA, followed by a wash in 0.1X SSC at 50EC for 45 min. Alternatively, filter washes can be performed in a solution containing 2 x SSC and 0.1% SDS, or 0.5 x SSC and 0.1% SDS, or 0.1 x SSC and 0.1% SDS at 68EC for 15 minute intervals. Following the wash steps, the hybridized probes are detectable by autoradiography. Other conditions of high stringency which may be used are well known in the art and as cited in Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., pp. 9.47-9.57; and Ausubel et al., 1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. are incorporated herein in their entirety.

(ii) By way of example and not limitation, procedures using conditions of intermediate stringency are as follows: Filters containing DNA are prehybridized, and then hybridized at a

temperature of 60EC in the presence of a 5 x SSC buffer and labeled probe. Subsequently, filters washes are performed in a solution containing 2x SSC at 50EC and the hybridized probes are detectable by autoradiography. Other conditions of intermediate stringency which may be used are well known in the art and as cited in Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., pp. 9.47-9.57; and Ausubel et al., 1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. are incorporated herein in their entirety.

d) a nucleotide sequence complementary to the sequence SEQ ID No. 1 or complementary to a nucleotide sequence as defined in a), b) or c) and a nucleotide sequence of their corresponding RNA;

e) a nucleotide sequence of a representative fragment of the sequence SEQ ID No. 1, or of a representative fragment of the nucleotide sequence as defined in a), b), c) or d);

f) a nucleotide sequence comprising a sequence as defined in a), b), c), d) or e);

g) a nucleotide sequence capable of being obtained from a nucleotide sequence as defined in a), b), c), d), e) or f); and

h) a modified nucleotide sequence of a nucleotide sequence as defined in a), b), c), d), e), f) or g).

Nucleotide sequence, polynucleotide or nucleic acid are understood to mean, according to the present invention, either a double-stranded DNA, a single-stranded DNA or products of transcription of the said DNAs.

It should be understood that the present invention does not relate to the genomic nucleotide sequences of *Chlamydia pneumoniae* taken in their natural environment, that is to say in the natural state. They are sequences which may have been isolated, purified or partially purified, by separation methods such as, for example, ion-exchange chromatography, molecular size exclusion chromatography or affinity chromatography, or alternatively fractionation techniques based on solubility in various solvents, or by genetic engineering methods such as amplification, cloning or subcloning, it being possible for the sequences of the invention to be carried by vectors.

The nucleotide sequence SEQ ID No. 1 was obtained by sequencing the *Chlamydia pneumoniae* genome by the method of directed sequencing after fluorescent automated sequencing of the inserts of clones and assembling of these sequences of nucleotide fragments (inserts) by means of softwares (cf. Examples). In spite of the high precision of the sequence SEQ ID No. 1, it is possible that it does not perfectly, 100% represent the nucleotide sequence of the *Chlamydia pneumoniae* genome and that a few rare sequencing errors or uncertainties still remain in the sequence SEQ ID No. 1. In the present invention, the presence of an uncertainty for an amino acid is designated by "Xaa" and that for a nucleotide is designated by "N" in the sequence listing below. These few rare errors or uncertainties could be easily detected and corrected by persons skilled in the art using the entire chromosome and/or its representative fragments according to the invention and standard

amplification, cloning and sequencing methods, it being possible for the sequences obtained to be easily compared, in particular by means of a computer software and using computer-readable media for recording the sequences according to the invention as described, for example, below. After correcting these possible rare errors or uncertainties, the corrected nucleotide sequence obtained would still exhibit at least 99.9% identity with the sequence SEQ ID No. 1. Such rare sequencing uncertainties are not present within the DNA contained within ATCC Deposit No. __ or ___, and whatever rare sequence uncertainties that exist within SEQ ID No. 1 can routinely be corrected utilizing the DNA of the ATCC deposits.

Homologous nucleotide sequence for the purposes of the present invention is understood to mean a nucleotide sequence having a percentage identity with the bases of the nucleotide sequence SEQ ID No. 1 of at least 80%, preferably 90% and 95%, this percentage being purely statistical and it being possible for the differences between the two nucleotide sequences to be distributed randomly and over their entire length. The said homologous sequences exhibiting a percentage identity with the bases of the nucleotide sequence SEQ ID No. 1 of at least 80%, preferably 90% and 95%, may comprise, for example, the sequences corresponding to the genomic sequence or to the sequences of its representative fragments of a bacterium belonging to the Chlamydia family, including the species *Chlamydia trachomatis*, *Chlamydia psittaci* and *Chlamydia pecorum* mentioned above, as well as the sequences corresponding to the genomic sequence or to the sequences of its representative fragments of a bacterium belonging to the variants of the species *Chlamydia pneumoniae*. In the present invention, the terms family and genus are mutually interchangeable, the terms variant, serotype, strain and subspecies are also mutually interchangeable. These homologous sequences may thus correspond to variations linked to mutations within the same species or between species and may correspond in particular to truncations, substitutions, deletions and/or additions of at least one nucleotide. The said homologous sequences may also correspond to variations linked to the degeneracy of the genetic code or to a bias in the genetic code which is specific to the family, to the species or to the variant and which are likely to be present in *Chlamydia*.

Protein and/or nucleic acid sequence homologies may be evaluated using any of the variety of sequence comparison algorithms and programs known in the art. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, and CLUSTALW (Pearson and Lipman, 1988, *Proc. Natl. Acad. Sci. USA 85(8)*:2444-2448; Altschul *et al.*, 1990, *J. Mol. Biol. 215(3)*:403-410; Thompson *et al.*, 1994, *Nucleic Acids Res. 22(2)*:4673-4680; Higgins *et al.*, 1996, *Methods Enzymol. 266*:383-402; Altschul *et al.*, 1990, *J. Mol. Biol. 215(3)*:403-410; Altschul *et al.*, 1993, *Nature Genetics 3*:266-272).

In a particularly preferred embodiment, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool ("BLAST") which is well known in the art (see, *e.g.*, Karlin and Altschul, 1990, *Proc. Natl. Acad. Sci. USA 87*:2267-2268; Altschul *et al.*, 1990, *J. Mol. Biol. 215*:403-410; Altschul *et al.*, 1993, *Nature Genetics 3*:266-272; Altschul *et al.*, 1997,

Nuc. Acids Res. 25:3389-3402). In particular, five specific BLAST programs are used to perform the following task:

(1)BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;

(2)BLASTN compares a nucleotide query sequence against a nucleotide sequence database;

(3)BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database;

(4)TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and

(5)TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (*i.e.*, aligned) by means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet *et al.*, 1992, *Science 256*:1443-1445; Henikoff and Henikoff, 1993, *Proteins 17*:49-61). Less preferably, the PAM or PAM250 matrices may also be used (see, *e.g.*, Schwartz and Dayhoff, eds., 1978, *Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure*, Washington: National Biomedical Research Foundation)

The BLAST programs evaluate the statistical significance of all high-scoring segment pairs identified, and preferably selects those segments which satisfy a user-specified threshold of significance, such as a user-specified percent homology. Preferably, the statistical significance of a high-scoring segment pair is evaluated using the statistical significance formula of Karlin (see, *e.g.*, Karlin and Altschul, 1990, *Proc. Natl. Acad. Sci. USA 87*:2267-2268).

Nucleotide sequence complementary to a sequence of the invention is understood to mean any DNA whose nucleotides are complementary to those of the sequence of the invention, and whose orientation is reversed (antiparallel sequence).

The present invention further comprises fragments of the sequences of a) through f), above. Representative fragments of the sequences according to the invention will be understood to mean any nucleotide fragment having at least 8 successive nucleotides, preferably at least 12 successive nucleotides, and still more preferably at least 15 or at least 20 successive nucleotides of the sequence from which it is derived. It is understood that such fragments refer only to portions of SEQ ID No. 1 that are not currently listed in a publicly available database.

Among these representative fragments, those capable of hybridizing under stringent conditions with a nucleotide sequence according to the invention are preferred. Hybridization under

stringent conditions means that the temperature and ionic strength conditions are chosen such that they allow hybridization to be maintained between two complementary DNA fragments.

By way of illustration, high stringency conditions for the hybridization step for the purposes of defining the nucleotide fragments described above, are advantageously the following.

The hybridization is carried out at a preferred temperature of 65EC in the presence of SSC buffer, 1 × SSC corresponding to 0.15 M NaCl and 0.05 M Na citrate. The washing steps may be, for example, the following:

2 × SSC, 0.1% SDS at room temperature followed by three washes with 1 × SSC, 0.1% SDS; 0.5 × SSC, 0.1% SDS; 0.1 × SSC, 0.1% SDS at 68EC for 15 minutes.

Intermediate stringency conditions, using, for example, a temperature of 60EC in the presence of a 5 × SSC buffer, or of low stringency, for example a temperature of 50EC in the presence of a 5 × SSC buffer, respectively require a lower overall complementarity for the hybridization between the two sequences.

The stringent hybridization conditions described above for a polynucleotide of about 300 bases in size will be adapted by persons skilled in the art for larger- or smaller-sized oligonucleotides, according to the teaching of Sambrook et al., 1989.

Among the representative fragments according to the invention, those which can be used as primer or probe in methods which make it possible to obtain homologous sequences or their representative fragments according to the invention, or to reconstitute a genomic fragment found to be incomplete in the sequence SEQ ID No. 1 or carrying an error or an uncertainty, are also preferred, these methods, such as the polymerase chain reaction (PCR), cloning and sequencing of nucleic acid being well known to persons skilled in the art. These homologous nucleotide sequences corresponding to mutations or to inter- or intra-species variations, as well as the complete genomic sequence or one of its representative fragments capable of being reconstituted, of course form part of the invention.

Among the said representative fragments, those which can be used as primer or probe in methods allowing diagnosis of the presence of *Chlamydia pneumoniae* or one of its associated microorganisms as defined below are also preferred.

The representative fragments capable of modulating, regulating, inhibiting or inducing the expression of a gene of *Chlamydia pneumoniae* or one of its associated microorganisms, and/or capable of modulating the replication cycle of *Chlamydia pneumoniae* or one of its associated microorganisms in the host cell and/or organism, are also preferred. Replication cycle is intended to designate invasion, multiplication, intracellular localization, in particular retention in the vacuole and inhibition of the process of fusion to the lysosome, and propagation of *Chlamydia pneumoniae* or one of its associated microorganisms from host cells to host cells.

Among the said representative fragments, those corresponding to nucleotide sequences corresponding to open reading frames, called ORF sequences (ORF for open reading frame), and

encoding polypeptides, such as for example, but without being limited thereto, the ORF sequences which will be later described, are finally preferred.

The representative fragments according to the invention may be obtained, for example, by specific amplification, such as PCR, or after digestion, with appropriate restriction enzymes, of nucleotide sequences according to the invention; these methods are in particular described in the manual by Sambrook et al., 1989. The said representative fragments may also be obtained by chemical synthesis when they are not too large in size and according to methods well known to persons skilled in the art. For example, such fragments can be obtained by isolating fragments of the genomic DNA of ATCC Deposit No. ___ or a clone insert present at this ATCC Deposit No. ___.

The representative fragments according to the invention may be used, for example, as primer, to reconstitute some of the said representative fragments, in particular those in which a portion of the sequence is likely to be missing or imperfect, by methods well known to persons skilled in the art such as amplification, cloning or sequencing techniques.

Modified nucleotide sequence will be understood to mean any nucleotide sequence obtained by mutagenesis according to techniques well known to persons skilled in the art, and exhibiting modifications in relation to the normal sequences, for example mutations in the regulatory and/or promoter sequences for the expression of a polypeptide, in particular leading to a modification of the level of expression of the said polypeptide or to a modulation of the replicative cycle.

Modified nucleotide sequence will also be understood to mean any nucleotide sequence encoding a modified polypeptide as defined below.

The subject of the present invention also includes *Chlamydia pneumoniae* nucleotide sequences characterized in that they are chosen from a nucleotide sequence of an open reading frame (ORF), that is, the ORF2 to ORF1297 sequences.

The ORF2 to ORF1297 nucleotide sequences are defined in Tables 1 and 2, *infra*, by their position on the sequence SEQ ID No. 1. For example, the ORF2 sequence is defined by the nucleotide sequence between the nucleotides at position 42 and 794 on the sequence SEQ ID No. 1, ends included. ORF2 to ORF1297 have been identified via homology analyses as well as via analyses of potential ORF start sites, as discussed in the examples below. It is to be understood that each identified ORF of the invention comprises a nucleotide sequence that spans the contiguous nucleotide sequence from the ORF stop codon immediately 3' to the stop codon of the preceding ORF and through the 5' codon to the next stop codon of SEQ ID No.:1 in-frame to the ORF nucleotide sequence. Table 2, *infra*, lists the beginning, end and potential start site of each of ORFs 1-1297. In one embodiment, the ORF comprises the contiguous nucleotide sequence spanning from the potential ORF start site downstream (that is, 3') to the ORF stop codon (or the ORF codon immediately adjacent to and upstream of the ORF stop codon). ORF2 to ORF1297 encode the polypeptides of SEQ ID No. 2 to SEQ ID No. 1291 and of SEQ ID No. 6844 to SEQ ID No. 6849, respectively.

Upon introduction of minor frameshifts, certain individual ORFs can comprise larger

"combined" ORFs. A list of such putative "combined" ORFs is shown in Table 3, below. For example, a combined ORF can comprise ORF 25, ORF 26 and ORF 27, including intervening in-frame, nucleotide sequences. The order of ORFs (5' to 3'), within each "combined" ORF is as listed. It is to be understood that when ORF2 to ORF1297 are referred to herein, such reference is also meant to include "combined" ORFs. Polypeptide sequences encoded by such "combined" ORFs are also part of the present invention.

Table 3

ORF 25, ORF 26, ORF 27;
ORF 28, ORF 29, ORF 30;
ORF 31, ORF 32;
ORF 33, ORF 35;
ORF 466, ORF 467;
ORF 468, ORF 469;
ORF 477, ORF 476, ORF 474;
ORF 480, ORF 482;
ORF 483, ORF 485, ORF 486, ORF 500;
ORF 503, ORF 504, ORF 505;
ORF 506, ORF 507;
ORF 1211, ORF 647;
ORF 1286, ORF 1039;
ORF 691, ORF 690;
ORF 105, ORF 106;
ORF 170, ORF 171;ORF 394, ORF 393;
ORF 453, ORF 452, ORF 451;
ORF 526, ORF 525;
ORF 757, ORF 756, ORF 755;
ORF 856, ORF 855;
ORF 958, ORF 957;
ORF 915, ORF 914, ORF 913;
ORF 543, ORF 544;
ORF 1266, ORF 380;
ORF 745, ORF 744;
ORF 777, ORF 776;
ORF 343, ORF 1297, and representative fragments.

Table 1 also depicts the results of homology searches that compared the sequences of the

polypeptides encoded by each of the ORFs to sequences present in public published databases. It is understood that those polypeptides listed in Table 1 as exhibiting greater than about 95% identity to a polypeptide present in a publicly disclosed database are not considered part of the present invention; likewise in this embodiment, those nucleotide sequences encoding such polypeptides are not considered part of the invention. In another embodiment, it is understood that those polypeptides listed in Table 1 as exhibiting greater than about 99% identity to a polypeptide present in a publicly disclosed database are not considered part of the invention; likewise, in this embodiment, those nucleotide sequences encoding such polypeptides are not considered part of the invention.

The invention also relates to the nucleotide sequences characterized in that they comprise a nucleotide sequence chosen from:

a) an ORF2 to ORF1297, a "combined" ORF nucleotide sequence, the nucleotide sequence of the genomic DNA contained within ATCC Deposit No. _______ or the nucleotide sequence of a clone insert in ATCC Deposit No. _______ according to the invention;

b) a homologous nucleotide sequence exhibiting at least 80% identity across an entire ORF2 to ORF1297 nucleotide sequence according to the invention or as defined in a);

c) a polynucleotide sequence that hybridizes to ORF2 to ORF1297 under conditions of high or intermediate stringency as described below:

(i) By way of example and not limitation, procedures using conditions of high stringency are as follows: Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65EC in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 μg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65EC, the preferred hybridization temperature, in prehybridization mixture containing 100 μg/ml denatured salmon sperm DNA and 5-20 X 10^6 cpm of ^{32}P-labeled probe. Alternatively, the hybridization step can be performed at 65EC in the presence of SSC buffer, 1 x SSC corresponding to 0.15M NaCl and 0.05 M Na citrate. Subsequently, filter washes can be done at 37EC for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA, followed by a wash in 0.1X SSC at 50EC for 45 min. Alternatively, filter washes can be performed in a solution containing 2 x SSC and 0.1% SDS, or 0.5 x SSC and 0.1% SDS, or 0.1 x SSC and 0.1% SDS at 68EC for 15 minute intervals. Following the wash steps, the hybridized probes are detectable by autoradiography. Other conditions of high stringency which may be used are well known in the art and as cited in Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., pp. 9.47-9.57; and Ausubel et al., 1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. are incorporated herein in their entirety. Preferably, such sequences encode a homolog of a polypeptide encoded by one of ORF2 to ORF1297. In one embodiment, such sequences encode a *Chlamydia pneumoniae* polypeptide.

(ii) By way of example and not limitation, procedures using conditions of intermediate

stringency are as follows: Filters containing DNA are prehybridized, and then hybridized at a temperature of 60EC in the presence of a 5 x SSC buffer and labeled probe. Subsequently, filters washes are performed in a solution containing 2x SSC at 50EC and the hybridized probes are detectable by autoradiography. Other conditions of intermediate stringency which may be used are well known in the art and as cited in Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., pp. 9.47-9.57; and Ausubel et al., 1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. are incorporated herein in their entirety. Preferably, such sequences encode a homolog of a polypeptide encoded by one of ORF2 to ORF1297. In one embodiment, such sequences encode a *Chlamydia pneumoniae* polypeptide.

d) complementary or RNA nucleotide sequence corresponding to an ORF2 to ORF1297 sequence according to the invention or as defined in a), b) or c);

e) a nucleotide sequence of a representative fragment of an ORF2 to ORF1297 sequence according to the invention or of a sequence as defined in a), b), c) or d);

f) a nucleotide sequence capable of being obtained from an ORF2 to ORF1297 sequence according to the invention or as defined in a), b), c), d) or e); and

g) a modified nucleotide sequence of an ORF2 to ORF1297 sequence according to the invention or as defined in a), b), c), d), e) or f);

As regards the homology with the ORF2 to ORF1297 nucleotide sequences, the homologous sequences exhibiting a percentage identity with the bases of one of the ORF2 to ORF1297 nucleotide sequences of at least 80%, preferably 90% and 95%, are preferred. Such homologous sequences are identified routinely via, for example, the algorithms described above and in the examples below. The said homologous sequences correspond to the homologous sequences as defined above and may comprise, for example, the sequences corresponding to the ORF sequences of a bacterium belonging to the Chlamydia family, including the species *Chlamydia trachomatis*, *Chlamydia psittaci* and *Chlamydia pecorum* mentioned above, as well as the sequences corresponding to the ORF sequences of a bacterium belonging to the variants of the species *Chlamydia pneumoniae*. These homologous sequences may likewise correspond to variations linked to mutations within the same species or between species and may correspond in particular to truncations, substitutions, deletions and/or additions of at least one nucleotide. The said homologous sequences may also correspond to variations linked to the degeneracy of the genetic code or to a bias in the genetic code which is specific to the family, to the species or to the variant and which are likely to be present in *Chlamydia*.

The invention comprises polypeptides encoded by a nucleotide sequence according to the invention, preferably by a representative fragment of the sequence SEQ ID No. 1 and corresponding to an ORF sequence, in particular the *Chlamydia pneumoniae* polypeptides, characterized in that they are chosen from the sequences SEQ ID No. 2 to SEQ ID No. 1291 or SEQ ID No. 6844 to SEQ ID No.

6849 and representative fragments thereof. However, the invention is not limited to polypeptides encoded by ORFs in SEQ ID No. 1 and its corresponding ORF sequences, but encompasses polypeptides of strain variants, polymorphisms, allelic variants, and mutants.

Thus, the invention also comprises the polypeptides characterized in that they comprise a polypeptide chosen from:

a) a polypeptide encoded by a polynucleotide sequence in SEQ ID No. 1 (*e.g.*, any polypeptide encoded by a polynucleotide sequence corresponding to ORF2 to ORF1297 and/or representative fragments thereof) according to the invention;

b) a polypeptide homologous to a polypeptide according to the invention, or as defined in a);

c) a polypeptide encoded by a polynucleotide sequence that hybridizes to SEQ ID No. 1 or ORF2 to ORF1297 under high or intermediate stringency as described below:

(i) By way of example and not limitation, procedures using conditions of high stringency are as follows: Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65EC in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 μg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65EC, the preferred hybridization temperature, in prehybridization mixture containing 100 μg/ml denatured salmon sperm DNA and 5-20 X 10^6 cpm of ^{32}P-labeled probe. Alternatively, the hybridization step can be performed at 65EC in the presence of SSC buffer, 1 x SSC corresponding to 0.15M NaCl and 0.05 M Na citrate. Subsequently, filter washes can be done at 37EC for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA, followed by a wash in 0.1X SSC at 50EC for 45 min. Alternatively, filter washes can be performed in a solution containing 2 x SSC and 0.1% SDS, or 0.5 x SSC and 0.1% SDS, or 0.1 x SSC and 0.1% SDS at 68EC for 15 minute intervals. Following the wash steps, the hybridized probes are detectable by autoradiography. Other conditions of high stringency which may be used are well known in the art and as cited in Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., pp. 9.47-9.57; and Ausubel et al., 1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. are incorporated herein in their entirety. Preferably such polypeptide represents a homolog of a polypeptide encoded by ORF2 to ORF1297. Preferably, such sequences encode a homolog of a polypeptide encoded by one of ORF2 to ORF1297. In one embodiment, such sequences encode a *Chlamydia pneumoniae* polypeptide.

(ii) By way of example and not limitation, procedures using conditions of intermediate stringency are as follows: Filters containing DNA are prehybridized, and then hybridized at a temperature of 60EC in the presence of a 5 x SSC buffer and labeled probe. Subsequently, filters washes are performed in a solution containing 2x SSC at 50EC and the hybridized probes are detectable by autoradiography. Other conditions of intermediate stringency which may be used are well known in the art and as cited in Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual,

Second Edition, Cold Spring Harbor Press, N.Y., pp. 9.47-9.57; and Ausubel et al., 1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. are incorporated herein in their entirety. Preferably, such sequences encode a homolog of a polypeptide encoded by one of ORF2 to ORF1297. In one embodiment, such sequences encode a *Chlamydia pneumoniae* polypeptide.

d) a fragment of at least 5 amino acids of a polypeptide according to the invention, or as defined in a), b) or c);

e) a biologically active fragment of a polypeptide according to the invention, or as defined in a), b), c) or d); and

f) a modified polypeptide of a polypeptide according to the invention, as defined in a), b), c),d) or e).

In the present description, the terms polypeptide, peptide and protein are interchangeable.

It should be understood that the invention does not relate to the polypeptides in natural form, that is to say that they are not taken in their natural environment but that they may have been isolated or obtained by purification from natural sources, or alternatively obtained by genetic recombination, or else by chemical synthesis and that they may, in this case, comprise nonnatural amino acids, as will be described below.

Homologous polypeptide will be understood to designate the polypeptides exhibiting, in relation to the natural polypeptide, certain modifications such as in particular a deletion, addition or substitution of at least one amino acid, a truncation, an extension, a chimeric fusion, and/or a mutation, or polypeptides exhibiting post-translational modifications. Among the homologous polypeptides, those whose amino acid sequence exhibits at least 80%, preferably 90%, homology or identity with the amino acid sequences of the polypeptides according to the invention are preferred. In the case of a substitution, one or more consecutive or nonconsecutive amino acids are replaced by "equivalent" amino acids. The expression "equivalent" amino acid is intended here to designate any amino acid capable of being substituted for one of the amino acids in the basic structure without, however, essentially modifying the biological activities of the corresponding peptides and as will be defined later.

Protein and/or nucleic acid sequence homologies may be evaluated using any of the variety of sequence comparison algorithms and programs known in the art. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, and CLUSTALW (Pearson and Lipman, 1988, *Proc. Natl. Acad. Sci. USA 85(8)*:2444-2448; Altschul *et al.*, 1990, *J. Mol. Biol. 215(3)*:403-410; Thompson *et al.*, 1994, *Nucleic Acids Res. 22(2)*:4673-4680; Higgins *et al.*, 1996, *Methods Enzymol. 266*:383-402; Altschul *et al.*, 1990, *J. Mol. Biol. 215(3)*:403-410; Altschul *et al.*, 1993, *Nature Genetics 3*:266-272).

In a particularly preferred embodiment, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool ("BLAST") which is well know in the art (see,

e.g., Karlin and Altschul, 1990, *Proc. Natl. Acad. Sci. USA 87*:2267-2268; Altschul *et al.*, 1990, *J. Mol. Biol. 215*:403-410; Altschul *et al.*, 1993, *Nature Genetics 3*:266-272; Altschul *et al.*, 1997, *Nuc. Acids Res. 25*:3389-3402). In particular, five specific BLAST programs are used to perform the following task:

(1)BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;

(2)BLASTN compares a nucleotide query sequence against a nucleotide sequence database;

(3)BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database;

(4)TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and

(5)TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (*i.e.*, aligned) by means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet *et al.*, 1992, *Science 256*:1443-1445; Henikoff and Henikoff, 1993, *Proteins 17*:49-61). Less preferably, the PAM or PAM250 matrices may also be used (see, *e.g.*, Schwartz and Dayhoff, eds., 1978, *Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure*, Washington: National Biomedical Research Foundation)

The BLAST programs evaluate the statistical significance of all high-scoring segment pairs identified, and preferably selects those segments which satisfy a user-specified threshold of significance, such as a user-specified percent homology. Preferably, the statistical significance of a high-scoring segment pair is evaluated using the statistical significance formula of Karlin (see, *e.g.*, Karlin and Altschul, 1990, *Proc. Natl. Acad. Sci. USA 87*:2267-2268).

Equivalent amino acids may be determined either based on their structural homology with the amino acids for which they are substituted, or on results of comparative tests of biological activity between the various polypeptides which may be carried out.

By way of example, there may be mentioned the possibilities of substitutions which may be carried out without resulting in a substantial modification of the biological activity of the corresponding modified polypeptides; the replacements, for example, of leucine with valine or isoleucine, of aspartic acid with glutamic acid, of glutamine with asparagine, of arginine with lysine, and the like, the reverse substitutions naturally being feasible under the same conditions.

The homologous polypeptides also correspond to the polypeptides encoded by the

homologous nucleotide sequences as defined above and thus comprise in the present definition the mutated polypeptides or polypeptides corresponding to inter- or intra-species variations which may exist in *Chlamydia*, and which correspond in particular to truncations, substitutions, deletions and/or additions of at least one amino acid residue.

Biologically active fragment of a polypeptide according to the invention will be understood to designate in particular a polypeptide fragment, as defined below, exhibiting at least one of the characteristics of the polypeptides according to the invention, in particular in that it is:

- capable of eliciting an immune response directed against *Chlamydia pneumoniae*; and/or
- capable of being recognized by an antibody specific for a polypeptide according to the invention; and/or
- capable of binding to a polypeptide or to a nucleotide sequence of *Chlamydia pneumoniae*; and/or
- capable of modulating, regulating, inducing or inhibiting the expression of a gene of *Chlamydia pneumoniae* or one of its associated microorganisms, and/or capable of modulating the replication cycle of *Chlamydia pneumoniae* or one of its associated microorganisms in the host cell and/or organism; and/or
- capable of generally exerting an even partial physiological activity, such as for example a structural activity (cellular envelope, ribosome), an enzymatic (metabolic) activity, a transport activity, an activity in the secretion or in the virulence.

A polypeptide fragment according to the invention is understood to designate a polypeptide comprising a minimum of 5 amino acids, preferably 10 amino acids or preferably 15 amino acids. It is to be understood that such fragments refer only to portions of polypeptides encoded by ORF2 to ORF1297 that are not currently listed in a publicly available database.

The polypeptide fragments according to the invention may correspond to isolated or purified fragments which are naturally present in *Chlamydia pneumoniae* or which are secreted by *Chlamydia pneumoniae*, or may correspond to fragments capable of being obtained by cleaving the said polypeptide with a proteolytic enzyme, such as trypsin or chymotrypsin or collagenase, or with a chemical reagent, such as cyanogen bromide (CNBr) or alternatively by placing the said polypeptide in a highly acidic environment, for example at pH 2.5. Such polypeptide fragments may be equally well prepared by chemical synthesis, using hosts transformed with an expression vector according to the invention containing a nucleic acid allowing the expression of the said fragments, placed under the control of appropriate elements for regulation and/or expression.

"Modified polypeptide" of a polypeptide according to the invention is understood to designate a polypeptide obtained by genetic recombination or by chemical synthesis as will be described below, exhibiting at least one modification in relation to the normal sequence. These modifications may in particular affect amino acids responsible for a specificity or for the efficiency of the activity, or responsible for the structural conformation, for the charge or for the hydrophobicity, and for the capacity for multimerization and for membrane insertion of the polypeptide according to

the invention. It is thus possible to create polypeptides with an equivalent, an increased or a reduced activity, and with an equivalent, a narrower or a broader specificity. Among the modified polypeptides, there may be mentioned the polypeptides in which up to 5 amino acids may be modified, truncated at the N- or C-terminal end, or alternatively deleted, or else added.

As is indicated, the modifications of the polypeptide may have in particular the objective:

- of making it capable of modulating, regulating, inhibiting or inducing the expression of a gene of *Chlamydia*, in particular of *Chlamydia pneumoniae* and its variants, or one of its associated microorganisms, and/or capable of modulating the replication cycle of *Chlamydia*, in particular of *Chlamydia pneumoniae* and its variants, or one of its associated microorganisms, in the host cell and/or organism,
- of allowing its use in methods of biosynthesis or of biodegradation, or its incorporation into vaccine compositions,
- of modifying its bioavailability as a compound for therapeutic use.

The said modified polypeptides may also be used on any cell or microorganism for which the said modified polypeptides will be capable of modulating, regulating, inhibiting or inducing gene expression, or of modulating the growth or the replication cycle of the said cell or of the said microorganism. The methods allowing demonstration of the said modulations on eukaryotic or prokaryotic cells are well known to persons skilled in the art. The said cells or microorganisms will be chosen, in particular, from tumour cells or infectious microorganisms and the said modified polypeptides may be used for the prevention or treatment of pathologies linked to the presence of the said cells or of the said microorganisms. It is also clearly understood that the nucleotide sequences encoding the said modified polypeptides may be used for the said modulations, for example by the intermediacy of vectors according to the invention and which are described below, so as to prevent or to treat the said pathologies.

The above modified polypeptides may be obtained using combinatory chemistry, in which it is possible to systematically vary portions of the polypeptide before testing them on models, cell cultures or microorganisms for example, so as to select the compounds which are the most active or which exhibit the desired properties.

Chemical synthesis also has the advantage of being able to use:

- nonnatural amino acids, or
- nonpeptide bonds.

Accordingly, in order to extend the life of the polypeptides according to the invention, it may be advantageous to use nonnatural amino acids, for example in the D form, or alternatively amino acid analogues, in particular sulphur-containing forms for example.

Finally, the structure of the polypeptides according to the invention, its homologous or modified forms, as well as the corresponding fragments may be integrated into chemical structures of the polypeptide type and the like. Accordingly, it may be advantageous to provide at the N- and C-

terminal ends compounds which are not recognized by proteases.

Also forming part of the invention are the nucleotide sequences encoding a polypeptide according to the invention. Described below are ORF nucleotide sequences encoding polypeptides exhibiting particularly preferable characteristics. For each group of preferred ORFS described below, it is to be understood that in addition to the individual ORFs listed, in instances wherein such ORFS are present as part of "combined" ORFs, the "combined" ORFs are also to be included within the preferred group.

More particularly, the subject of the invention is nucleotide sequences, characterized in that they encode a polypeptide of the cellular envelope, preferably of the outer cellular envelope of *Chlamydia pneumoniae* or one of its representative fragments, such as for example the predominant proteins of the outer membrane, the adhesion proteins or the proteins entering into the composition of the Chlamydia wall. Among these sequences, the sequences comprising a nucleotide sequence chosen from the following sequences are most preferred:

ORF15; ORF25; ORF26; ORF27; ORF28; ORF29; ORF30; ORF31; ORF32; ORF33; ORF35; ORF68; ORF124; ORF275; ORF291; ORF294; ORF327; ORF342; ORF364; ORF374; ORF380; ORF414; ORF439; ORF466; ORF467; ORF468; ORF469; ORF470; ORF472; ORF474; ORF476; ORF477; ORF478; ORF479; ORF480; ORF482; ORF485; ORF500; ORF501; ORF503; ORF504; ORF505; ORF506; ORF520; ORF578; ORF580; ORF581; ORF595; ORF596; ORF597; ORF737; ORF830; ORF834; ORF836; ORF893; ORF917; ORF932; ORF976; ORF1035; ORF1045; ORF1090 and one of their representative fragments.

The structure of the cytoplasmic membranes and of the wall of bacteria is dependent on the associated proteins. The structure of the cytoplasmic membrane makes it impermeable to water, to water-soluble substances and to small-sized molecules (ions, small inorganic molecules, peptides or proteins). To enter into or to interfere with a cell or a bacterium, a ligand must establish a special relationship with a protein anchored in the cytoplasmic membrane (the receptor). These proteins which are anchored on the membrane play an important role in metabolism since they control the exchanges in the bacterium. These exchanges apply to molecules of interest for the bacterium (small molecules such as sugars and small peptides) as well as undesirable molecules for the bacterium such as antibiotics or heavy metals.

The double lipid layer structure of the membrane requires the proteins which are inserted therein to have hydrophobic domains of about twenty amino acids forming an alpha helix. Predominantly hydrophobic and potentially transmembrane regions may be predicted from the primary sequence of the proteins, itself deduced from the nucleotide sequence. The presence of one or more putative transmembrane domains raises the possibility for a protein to be associated with the cytoplasmic membrane and to be able to play an important metabolic role therein or alternatively for the protein thus exposed to be able to exhibit potentially protective epitopes.

If the proteins inserted into the membrane exhibit several transmembrane domains

capable of interacting with one another via electrostatic bonds, it then becomes possible for these proteins to form pores which go across the membrane which becomes permeable for a number of substances. It should be noted that proteins which do not have transmembrane domains may also be anchored by the intermediacy of fatty acids in the cytoplasmic membrane, it being possible for the breaking of the bond between the protein and its anchor in some cases to be responsible for the release of the peptide outside the bacterium.

Preferably, the invention relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* transmembrane polypeptide or one of its representative fragments, having between 1 and 3 transmembrane domains and in that they comprise a nucleotide sequence chosen from the following sequences:

ORF2; ORF3; ORF6; ORF9; ORF10; ORF11; ORF13; ORF14; ORF16; ORF18; ORF19; ORF20; ORF21; ORF22; ORF25; ORF27; ORF28; ORF29; ORF30; ORF31; ORF32; ORF33; ORF34; ORF35; ORF37; ORF39; ORF41; ORF42; ORF44; ORF45; ORF46; ORF47; ORF48; ORF49; ORF50; ORF53; ORF54; ORF56; ORF57; ORF59; ORF60; ORF61; ORF62; ORF63; ORF64; ORF65; ORF66; ORF69;; ORF72; ORF73; ORF74; ORF76; ORF77; ORF78; ORF79; ORF80; ORF82; ORF84; ORF85; ORF86; ORF88; ORF89; ORF90; ORF91; ORF92; ORF93; ORF95; ORF96; ORF98; ORF99; ORF100; ORF101; ORF102; ORF103; ORF104; ORF105; ORF106; ORF107; ORF108; ORF114; ORF117; ORF118; ORF122; ORF123; ORF124; ORF125; ORF129; ORF130; ORF131; ORF132; ORF133; ORF134; ORF135; ORF137; ORF138; ORF139; ORF140; ORF141; ORF142; ORF143; ORF145; ORF146; ORF147; ORF150; ORF151; ORF152; ORF156; ORF157; ORF158; ORF159; ORF160; ORF161; ORF162; ORF164; ORF166; ORF167; ORF170; ORF173; ORF175; ORF176; ORF178; ORF179; ORF180; ORF182; ORF183; ORF184; ORF185; ORF186; ORF187; ORF188; ORF189; ORF190; ORF191; ORF192; ORF194; ORF195; ORF196; ORF197; ORF198; ORF199; ORF200; ORF201; ORF202; ORF205; ORF207; ORF208; ORF209; ORF210; ORF212; ORF215; ORF219; ORF220; ORF224; ORF226; ORF227; ORF228; ORF231; ORF232; ORF233; ORF234; ORF235; ORF236; ORF238; ORF239; ORF240; ORF241; ORF242; ORF244; ORF247; ORF251; ORF252; ORF253; ORF255; ORF256; ORF257; ORF258; ORF260; ORF262; ORF263; ORF266; ORF267; ORF268; ORF269; ORF270; ORF273; ORF274; ORF276; ORF278; ORF279; ORF280; ORF281; ORF282; ORF283; ORF284; ORF286; ORF287; ORF289; ORF290; ORF291; ORF293; ORF294; ORF297; ORF304; ORF305; ORF307; ORF308; ORF309; ORF310; ORF311; ORF313; ORF314; ORF315; ORF316; ORF318; ORF319; ORF320; ORF321; ORF322; ORF323; ORF324; ORF325; ORF326; ORF331; ORF332; ORF336; ORF338; ORF339; ORF341; ORF344; ORF345; ORF346; ORF350; ORF352; ORF353; ORF356; ORF357; ORF358; ORF359; ORF360; ORF362; ORF365; ORF366; ORF367; ORF370; ORF372; ORF373; ORF376; ORF377; ORF378; ORF379; ORF381; ORF382; ORF383; ORF384; ORF385; ORF386; ORF387; ORF390; ORF392; ORF393; ORF394; ORF396; ORF398; ORF399; ORF400; ORF404; ORF408; ORF410; ORF411; ORF413; ORF416; ORF417; ORF418; ORF420; ORF422; ORF424; ORF427;

ORF428; ORF429; ORF430; ORF431; ORF433; ORF434; ORF437; ORF440; ORF441; ORF442; ORF443; ORF444; ORF445; ORF447; ORF450; ORF451; ORF452; ORF455; ORF456; ORF459; ORF460; ORF461; ORF462; ORF463; ORF464; ORF465; ORF467; ORF469; ORF471; ORF474; ORF475; ORF476; ORF477; ORF479; ORF482; ORF483; ORF484; ORF485; ORF486; ORF487; ORF488; ORF491; ORF493; ORF494; ORF497; ORF498; ORF499; ORF503; ORF508; ORF509; ORF510; ORF512; ORF514; ORF515; ORF516; ORF517; ORF518; ORF520; ORF521; ORF523; ORF525; ORF527; ORF528; ORF529; ORF530; ORF531; ORF533; ORF534; ORF535; ORF536; ORF537; ORF540; ORF541; ORF543; ORF544; ORF545; ORF546; ORF548; ORF549; ORF551; ORF553; ORF554; ORF555; ORF556; ORF557; ORF558; ORF559; ORF560; ORF562; ORF563; ORF564; ORF565; ORF566; ORF569; ORF571; ORF573; ORF576; ORF577; ORF581; ORF583; ORF584; ORF585; ORF586; ORF588; ORF591; ORF592; ORF594; ORF595; ORF596; ORF597; ORF599; ORF600; ORF603; ORF605; ORF608; ORF614; ORF615; ORF620; ORF621; ORF622; ORF623; ORF624; ORF625; ORF629; ORF630; ORF631; ORF633; ORF634; ORF637; ORF642; ORF644; ORF645; ORF647; ORF648; ORF652; ORF654; ORF655; ORF657; ORF658; ORF659; ORF660; ORF661; ORF664; ORF665; ORF666; ORF667; ORF670; ORF671; ORF672; ORF673; ORF674; ORF676; ORF679; ORF681; ORF684; ORF687; ORF688; ORF689; ORF690; ORF693; ORF694; ORF695; ORF696; ORF697; ORF698; ORF699; ORF700; ORF701; ORF703; ORF705; ORF706; ORF707; ORF708; ORF710; ORF712; ORF715; ORF716; ORF717; ORF718; ORF719; ORF721; ORF722; ORF723; ORF725; ORF726; ORF727; ORF728; ORF729; ORF730; ORF731; ORF733; ORF736; ORF737; ORF738; ORF740; ORF741; ORF742; ORF743; ORF747; ORF748; ORF750; ORF752; ORF754; ORF755; ORF756; ORF757; ORF759; ORF760; ORF761; ORF762; ORF763; ORF764; ORF765; ORF766; ORF767; ORF768; ORF772; ORF774; ORF775; ORF777; ORF781; ORF783; ORF788; ORF791; ORF792; ORF793; ORF794; ORF795; ORF796; ORF797; ORF798; ORF799; ORF802; ORF803; ORF806; ORF807; ORF808; ORF809; ORF810; ORF811; ORF813; ORF814; ORF815; ORF816; ORF817; ORF819; ORF820; ORF821; ORF823; ORF824; ORF827; ORF829; ORF830; ORF831; ORF833; ORF834; ORF835; ORF837; ORF844; ORF845; ORF846; ORF847; ORF848; ORF849; ORF850; ORF851; ORF852; ORF854; ORF855; ORF856; ORF857; ORF859; ORF860; ORF862; ORF865; ORF866; ORF868; ORF869; ORF870; ORF871; ORF872; ORF874; ORF877; ORF878; ORF879; ORF880; ORF881; ORF882; ORF884; ORF885; ORF888; ORF889; ORF890; ORF891; ORF892; ORF894; ORF895; ORF896; ORF897; ORF899; ORF900; ORF902; ORF903; ORF904; ORF905; ORF909; ORF910; ORF912; ORF913; ORF914; ORF915; ORF917; ORF918; ORF919; ORF921; ORF923; ORF924; ORF926; ORF927; ORF928; ORF929; ORF930; ORF931; ORF937; ORF938; ORF939; ORF941; ORF943; ORF948; ORF951; ORF952; ORF953; ORF958; ORF960; ORF963; ORF964; ORF965; ORF968; ORF970; ORF974; ORF975; ORF977; ORF979; ORF980; ORF981; ORF983; ORF984; ORF985; ORF987; ORF989; ORF992; ORF993; ORF997; ORF998; ORF999; ORF1001; ORF1002; ORF1004; ORF1005; ORF1009; ORF1013; ORF1014; ORF1015; ORF1016; ORF1019; ORF1021; ORF1023; ORF1024;

ORF1029; ORF1031; ORF1033; ORF1034; ORF1039; ORF1041; ORF1042; ORF1045; ORF1047; ORF1049; ORF1051; ORF1052; ORF1053; ORF1054; ORF1056; ORF1059; ORF1061; ORF1062; ORF1063; ORF1064; ORF1065; ORF1067; ORF1075; ORF1077; ORF1078; ORF1079; ORF1080; ORF1081; ORF1089; ORF1095; ORF1097; ORF1098; ORF1099; ORF1101; ORF1102; ORF1103; ORF1106; ORF1107; ORF1108; ORF1109; ORF1110; ORF1113; ORF1116; ORF1118; ORF1119; ORF1121; ORF1123; ORF1124; ORF1126; ORF1128; ORF1130; ORF1131; ORF1133; ORF1134; ORF1136; ORF1137 and one of their representative fragments.

Preferably, the invention relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* transmembrane polypeptide or one of its representative fragments, having between 4 and 6 transmembrane domains and in that they comprise a nucleotide sequence chosen from the following sequences:

ORF5; ORF7; ORF8; ORF15; ORF36; ORF38; ORF51; ORF55; ORF58; ORF67; ORF70; ORF81; ORF97; ORF110; ORF111; ORF115; ORF119; ORF126; ORF128; ORF148; ORF155; ORF163; ORF165; ORF168; ORF169; ORF171; ORF172; ORF174; ORF177; ORF181; ORF193; ORF203; ORF213; ORF214; ORF216; ORF217; ORF221; ORF222; ORF225; ORF229; ORF243; ORF246; ORF248; ORF254; ORF261; ORF285; ORF288; ORF292; ORF296; ORF298; ORF299; ORF301; ORF303; ORF317; ORF328; ORF329; ORF351; ORF354; ORF355; ORF364; ORF371; ORF374; ORF375; ORF391; ORF395; ORF401; ORF403; ORF405; ORF409; ORF414; ORF419; ORF421; ORF423; ORF425; ORF438; ORF448; ORF453; ORF458; ORF466; ORF468; ORF470; ORF480; ORF489; ORF490; ORF496; ORF501; ORF504; ORF505; ORF506; ORF511; ORF513; ORF519; ORF526; ORF532; ORF538; ORF539; ORF547; ORF550; ORF561; ORF568; ORF570; ORF574; ORF578; ORF579; ORF580; ORF582; ORF589; ORF593; ORF598; ORF601; ORF604; ORF610; ORF613; ORF617; ORF626; ORF632; ORF635; ORF638; ORF640; ORF641; ORF646; ORF649; ORF650; ORF651; ORF686; ORF711; ORF724; ORF732; ORF734; ORF744; ORF745; ORF749; ORF751; ORF769; ORF770; ORF771; ORF773; ORF776; ORF779; ORF780; ORF785; ORF787; ORF789; ORF801; ORF805; ORF812; ORF822; ORF825; ORF826; ORF839; ORF841; ORF843; ORF853; ORF861; ORF875; ORF876; ORF886; ORF893; ORF898; ORF906; ORF907; ORF908; ORF920; ORF922; ORF925; ORF933; ORF935; ORF936; ORF944; ORF946; ORF947; ORF954; ORF959; ORF961; ORF966; ORF967; ORF972; ORF978; ORF995; ORF996; ORF1000; ORF1003; ORF1010; ORF1011; ORF1012; ORF1017; ORF1020; ORF1030; ORF1036; ORF1038; ORF1043; ORF1046; ORF1048; ORF1050; ORF1058; ORF1071; ORF1073; ORF1084; ORF1085; ORF1086; ORF1087; ORF1091; ORF1092; ORF1094; ORF1096; ORF1100; ORF1104; ORF1111; ORF1112; ORF1114; ORF1117; ORF1122; ORF1125 and one of their representative fragments.

Preferably, the invention also relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* transmembrane polypeptide or one of its representative fragments, having at least 7 transmembrane domains and in that they comprise a nucleotide sequence chosen from the following sequences:

ORF17; ORF52; ORF68; ORF83; ORF87; ORF109; ORF112; ORF113; ORF120; ORF121; ORF127; ORF153; ORF204; ORF211; ORF218; ORF223; ORF275; ORF277; ORF295; ORF300; ORF302; ORF306; ORF327; ORF335; ORF342; ORF343; ORF347; ORF349; ORF361; ORF363; ORF369; ORF380; ORF388; ORF389; ORF397; ORF415; ORF432; ORF439; ORF446; ORF449; ORF472; ORF478; ORF500; ORF522; ORF524; ORF567; ORF575; ORF602; ORF606; ORF609; ORF636; ORF639; ORF643; ORF653; ORF668; ORF692; ORF702; ORF704; ORF713; ORF720; ORF778; ORF784; ORF800; ORF836; ORF838; ORF842; ORF864; ORF867; ORF883; ORF901; ORF916; ORF932; ORF934; ORF940; ORF942; ORF950; ORF956; ORF971; ORF973; ORF976; ORF988; ORF994; ORF1018; ORF1028; ORF1035; ORF1037; ORF1044; ORF1055; ORF1057; ORF1068; ORF1069; ORF1070; ORF1072; ORF1082; ORF1088; ORF1105; ORF1132; ORF1135 and one of their representative fragments.

Preferably, the invention relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* surface exposed polypeptide (*e.g.*, an outer membrane protein) or one of its representative fragments, said nucleotide sequences comprising a nucleotide sequence chosen from the following sequences:

ORF 15, ORF 25, ORF 26, ORF 27, ORF 28, ORF 29, ORF 30, ORF 31, ORF 32, ORF 33, ORF 35, ORF 36, ORF 1257, ORF 280, ORF 291, ORF 314, ORF 354, ORF 380, ORF 1266, ORF 466, ORF 467, ORF 468, ORF 469, ORF 470, ORF 472, ORF 474, ORF 476, ORF 477, ORF 478, ORF 479, ORF 480, ORF 482, ORF 483, ORF 485, ORF 486, ORF 500, ORF 501, ORF 503, ORF 504, ORF 505, ORF 506, ORF 507, ORF 1268, ORF 1269, ORF 543, ORF 544, ORF 578, ORF 579, ORF 580, ORF 581, ORF 595, ORF 596, ORF 597, ORF 1271, ORF 633, ORF 637, ORF 699, ORF 706, ORF 737, ORF 744, ORF 1273, ORF 751, ORF 775, ORF 776, ORF 777, ORF 793, ORF 815, ORF 830, ORF 1221, ORF 849, ORF 851, ORF 852, ORF 874, ORF 891, ORF 922, ORF 940, ORF 1231, ORF 1281, ORF 1035, ORF 1079, ORF 1087, ORF 1108, and one of their representative fragments.

Preferably, the invention relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* lipoprotein or one of its representative fragments, said nucleotide sequences comprising a nucleotide sequence chosen from the following sequences:

ORF 3, ORF 10, ORF 11, ORF 16, ORF 1254, ORF 1255, ORF 38, ORF 1256, ORF 62, ORF 85, ORF 1258, ORF 115, ORF 1151, ORF 151, ORF 1259, ORF 173, ORF 1261, ORF 186, ORF 194, ORF 205, ORF 214, ORF 216, ORF 217, ORF 238, ORF 1177, ORF 280, ORF 291, ORF 317, ORF 327, ORF 354, ORF 364, ORF 367, ORF 414, ORF 432, ORF 1192, ORF 460, ORF 1267, ORF 1268, ORF 520, ORF 536, ORF 1270, ORF 576, ORF 597, ORF 603, ORF 609, ORF 637, ORF 1272, ORF 652, ORF 1213, ORF 699, ORF 705, ORF 706, ORF 708, ORF 711, ORF 727, ORF 1274, ORF 800, ORF 814, ORF 825, ORF 829, ORF 830, ORF 831, ORF 844, ORF 849, ORF 1275, ORF 1276, ORF 1277, ORF 872, ORF 878, ORF 880, ORF 891, ORF 892, ORF 1278, ORF 1279, ORF 1280, ORF 941, ORF 942, ORF 1282, ORF 1283, ORF 952, ORF 988, ORF 998, ORF 1009, ORF 1285, ORF

1235, ORF 1028, ORF 1056, ORF 1070, ORF 1287, ORF 1087, ORF 1288, ORF 1289, ORF 1098, ORF 1246, ORF 1291, ORF 1108, ORF 1109, ORF 1112, ORF 1133, and one of their representative fragments.

Preferably, the invention relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* polypeptide involved in lipopolysaccharide (LPS) biosynthesis, said nucleotide sequences comprising a nucleotide sequence chosen from the following sequences: ORF 316, ORF 564, ORF 610, ORF 647, ORF 1211, ORF 688, ORF 924, and one of their representative fragments.

Preferably the invention relates to additional LPS-related nucleotide sequences according to the invention, characterized in that they encode:

(a) a *Chlamydia pneumoniae* KDO (3-deoxy-D-manno-octulosonic acid)-related polypeptide or one of its representative fragments, said nucleotide sequences comprising a nucleotide sequence chosen from the following sequences: ORF 177, ORF 1156, ORF 245, ORF 767, and one of their representative fragments;

(b) a *Chlamydia pneumoniae* phosphomannomutase-related polypeptide or one of its representative fragments, said nucleotide sequences comprising a nucleotide sequence chosen from the following sequences: ORF 74, and one of its representative fragments;

(c) a *Chlamydia pneumoniae* phosphoglucomutase-related polypeptide or one of its representative fragments, said nucleotide sequences comprising a nucleotide sequence chosen from the following sequences: ORF 1286, ORF 1039, and one of their representative fragments; and

(d) a *Chlamydia pneumoniae* lipid A component-related polypeptide or one of its representative fragments, said nucleotide sequences comprising a nucleotide sequence chosen from the following sequences: ORF 689, ORF 690, ORF 691, ORF 1037, and one of their representative fragments.

Preferably, the invention relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* polypeptide containing RGD (Arg-Gly-Asp) attachment sites or one of its representative fragments.

(a) RGD-containing proteins that are outer membrane proteins, are more likely to play a role in cell attachment. ORFs that encoded a protein containing an RGD sequence and also were classified as outer membrane proteins are ORF 468 and its representative fragments.

(b) An RGD-encoding ORF that showed homology to cds1, cds2, and copN type III virulence loci in *Chlamydia psittaci* (Hsia, R. et al. (1997), Type III secretion genes identity a putative virulence locus of Chlamydia. Molecular Microbiology 25:351-359) is ORF 350, and its representative fragments.

(c) The outer membrane of Chlamydia is made of cysteine-rich proteins that form a network of both intra and inter molecular disulfide links. This contributes to the integrity of the membrane since Chlamydia lacks the peptidoglycan layer that other gram-negative bacteria have. Cysteine-rich proteins that have the RGD sequence are also considered to be potential vaccine candidates. Cysteine-rich proteins were defined as proteins that had more than 3.0% cysteine in their primary amino acid sequence, above the mean genomic ORF cysteine content. The corresponding ORFs are: ORF 1290, ORF 1294, ORF 1296, and one of their representative fragments.

(d) The outer membrane of Chlamydia may also contain small proteins that have cysteines in their N– and C-terminus that may contribute to the network formed by disulfide linkages. These proteins may be anchored in the outer membrane via their N-terminus and may have their C-terminus exposed, which then can interact with the host cells. Alternatively, these proteins may be anchored in the outer membrane via both N-and C-terminus and may have regions in the middle that may be exposed which can in turn interact with the host cells. ORFs encoding polypeptides that contain cysteines in their first 30 amino acids and also contain an RGD sequence are: ORF 105, ORF 106, ORF 114, ORF 170, ORF 171, ORF 1264, ORF 268, ORF 1265, ORF 350, ORF 393, ORF 394, ORF 451, ORF 452, ORF 453, ORF 473, ORF 499, ORF 515, ORF 519, ORF 525, ORF 526, ORF 538, ORF 611, ORF 645, ORF 686, ORF 700, ORF 746, ORF 755, ORF 756, ORF 757, ORF 789, ORF 814, ORF 855, ORF 856, ORF 878, ORF 957, ORF 958, ORF 989, ORF 1290, and one of their representative fragments.

(e) RGD-containing ORFs homologous to RGD-containing ORFs from *Chlamydia trachomatis* are:
ORF 114, ORF 468, ORF 755, ORF 756, ORF 757, ORF 855, ORF 856, ORF 905, ORF 913, ORF 914, ORF 915, and one of their representative fragments.

Preferably, the invention relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* Type III or other, non-type III secreted polypeptide or one of its representative fragments, said nucleotide sequences comprising a nucleotide sequence chosen from the following sequences:
ORF 25, ORF 28, ORF 29, ORF 33, ORF 308, ORF 309, ORF 343, ORF 344, ORF 345, ORF 367, ORF 414, ORF 415, ORF 480, ORF 550, ORF 579, ORF 580, ORF 581, ORF 597, ORF 699, ORF 744, ORF 751, ORF 776, ORF 866, ORF 874, ORF 883, ORF 884, ORF 888, ORF 891, ORF 1293,

and one of their representative fragments.

Preferably, the invention relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* cell wall anchored surface polypeptide or one of its representative fragments, said nucleotide sequences comprising a nucleotide sequence chosen from the following sequences: ORF 267, ORF 271, ORF 419, ORF 590, ORF 932, ORF 1292, ORF 1295, and one of their representative fragments.

Preferably, the invention relates to the nucleotide sequences according to the invention, characterized in that they encode *Chlamydia pneumoniae* polypeptides not found in *Chlamydia trachomatis* (Blastp. P>e^{-10}), said nucleotide sequences comprising a nucleotide sequence chosen from the following sequences: ORF 7, ORF 8, ORF 9, ORF 16, ORF 17, ORF 18, ORF 19, ORF 20, ORF 21, ORF 22, ORF 1254, ORF 23, ORF 1255, ORF 24, ORF 1139, ORF 1140, ORF 46, ORF 47, ORF 51, ORF 60, ORF 1256, ORF 61, ORF 62, ORF 63, ORF 64, ORF 1257, ORF 65, ORF 66, ORF 67, ORF 68, ORF 1143, ORF 1145, ORF 83, ORF 84, ORF 1146, ORF 85, ORF 86, ORF 87, ORF 1258, ORF 116, ORF 117, ORF 125, ORF 1148, ORF 143, ORF 1150, ORF 1151, ORF 144, ORF 145, ORF 147, ORF 148, ORF 149, ORF 150, ORF 152, ORF 1259, ORF 162, ORF 166, ORF 1154, ORF 167, ORF 1261, ORF 1156, ORF 1157, ORF 178, ORF 179, ORF 1158, ORF 182, ORF 183, ORF 184, ORF 185, ORF 1159, ORF 186, ORF 1160, ORF 187, ORF 188, ORF 189, ORF 190, ORF 1161, ORF 1162, ORF 191, ORF 192, ORF 194, ORF 195, ORF 1163, ORF 196, ORF 201, ORF 202, ORF 209, ORF 212, ORF 221, ORF 224, ORF 1167, ORF 226, ORF 227, ORF 228, ORF 229, ORF 230, ORF 231, ORF 232, ORF 1169, ORF 1170, ORF 1171, ORF 234, ORF 235, ORF 236, ORF 1172, ORF 243, ORF 251, ORF 252, ORF 1176, ORF 253, ORF 255, ORF 254, ORF 256, ORF 1177, ORF 1178, ORF 262, ORF 263, ORF 1264, ORF 278, ORF 279, ORF 1180, ORF 280, ORF 290, ORF 291, ORF 292, ORF 296, ORF 1181, ORF 297, ORF 298, ORF 300, ORF 1265, ORF 322, ORF 324, ORF 325, ORF 370, ORF 1186, ORF 371, ORF 372, ORF 1187, ORF 373, ORF 378, ORF 1266, ORF 382, ORF 383, ORF 384, ORF 385, ORF 386, ORF 1188, ORF 1189, ORF 391, ORF 392, ORF 398, ORF 400, ORF 403, ORF 1191, ORF 423, ORF 435, ORF 445, ORF 450, ORF 1193, ORF 456, ORF 460, ORF 461, ORF 465, ORF 1196, ORF 471, ORF 473, ORF 475, ORF 481, ORF 484, ORF 487, ORF 488, ORF 489, ORF 490, ORF 491, ORF 492, ORF 493, ORF 494, ORF 495, ORF 496, ORF 497, ORF 498, ORF 499, ORF 502, ORF 1267, ORF 1268, ORF 508, ORF 510, ORF 509, ORF 512, ORF 515, ORF 519, ORF 1197, ORF 521, ORF 1198, ORF 522, ORF 524, ORF 528, ORF 534, ORF 537, ORF 1269, ORF 1270, ORF 548, ORF 551, ORF 557, ORF 1201, ORF 1203, ORF 562, ORF 566, ORF 593, ORF 595, ORF 600, ORF 1271, ORF 604, ORF 611, ORF 612, ORF 614, ORF 616, ORF 625, ORF 627, ORF 628, ORF 629, ORF 631, ORF 641, ORF 1272, ORF 648, ORF 1212, ORF 663, ORF 685, ORF 707, ORF 714, ORF 715, ORF 716, ORF 717, ORF 722, ORF 746, ORF 1273, ORF 761, ORF 764, ORF 770, ORF 1217, ORF 783, ORF 1274, ORF 803, ORF 815, ORF 1220, ORF 835, ORF 1221, ORF 844, ORF 845, ORF 846, ORF 847, ORF 848, ORF 849, ORF 850, ORF 851, ORF 1275, ORF 852, ORF 862, ORF 1276, ORF 1277, ORF 873, ORF 1223, ORF 892, ORF 919, ORF 1225,

ORF 1278, ORF 926, ORF 1228, ORF 1229, ORF 1230, ORF 1279, ORF 1281, ORF 1282, ORF 1283, ORF 948, ORF 950, ORF 949, ORF 951, ORF 980, ORF 982, ORF 1233, ORF 999, ORF 1000, ORF 1001, ORF 1002, ORF 1008, ORF 1285, ORF 1235, ORF 1016, ORF 1019, ORF 1027, ORF 1036, ORF 1241, ORF 1048, ORF 1049, ORF 1050, ORF 1053, ORF 1054, ORF 1064, ORF 1076, ORF 1091, ORF 1288, ORF 1093, ORF 1289, ORF 1101, ORF 1103, ORF 1245, ORF 1246, ORF 1247, ORF 1290, ORF 1291, ORF 1115, ORF 1116, ORF 1118, ORF 1120, ORF 1249, ORF 1121, ORF 1250, ORF 1126, ORF 1251, ORF 1127, ORF 1128, ORF 1130, ORF 1129, ORF 1131, ORF 1136, ORF 1253, ORF 1292, ORF 1294, ORF 1295, ORF 1296, and one of their representative fragments.

Preferably, the invention also relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* polypeptide or one of its representative fragments which is involved in the intermediate metabolism, in particular in the metabolism of sugars and/or of cofactors, such as for example triose phosphate isomerase or pyruvate kinase, and in that they comprise a nucleotide sequence chosen from the following sequences:

ORF2; ORF55; ORF56; ORF69; ORF75; ORF80; ORF100; ORF110; ORF114; ORF120; ORF121; ORF157; ORF160; ORF161; ORF172; ORF180; ORF181; ORF198; ORF200; ORF225; ORF248; ORF249; ORF276; ORF277; ORF318; ORF319; ORF320; ORF323; ORF331; ORF347; ORF375; ORF376; ORF381; ORF393; ORF394; ORF395; ORF396; ORF409; ORF446; ORF447; ORF448; ORF449; ORF513; ORF516; ORF571; ORF647; ORF662; ORF697; ORF718; ORF793; ORF794; ORF808; ORF809; ORF838; ORF839; ORF840; ORF853; ORF854; ORF918; ORF923; ORF929; ORF931; ORF938; ORF939; ORF958; ORF959; ORF960; ORF966; ORF995; ORF1021; ORF1040; ORF1041; ORF1042; ORF1085; ORF1100; ORF1102; ORF1117; ORF1118; ORF1119; ORF1120; ORF1135 and one of their representative fragments.

Preferably, the invention also relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* polypeptide or one of its representative fragments which is involved in the intermediate metabolism of nucleotides or nucleic acids, such as for example CTP synthetase or GMP synthetase, and in that they comprise a nucleotide sequence chosen from the following sequences:

ORF77; ORF78; ORF138; ORF189; ORF190; ORF233; ORF246; ORF338; ORF412; ORF421; ORF438; ORF607; ORF648; ORF657; ORF740; ORF783; ORF967; ORF989; ORF990; ORF992; ORF1011; ORF1058; ORF1059; ORF1073; ORF1074 and one of their representative fragments.

Preferably, the invention also relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* polypeptide or one of its representative fragments which is involved in the metabolism of nucleic acids, such as for example DNA polymerases or DNA topoisomerases, and in that they comprise a nucleotide sequence chosen from the following sequences:

ORF14; ORF59; ORF70; ORF71; ORF97; ORF113; ORF137; ORF141; ORF169; ORF285; ORF287;

ORF288; ORF313; ORF326; ORF358; ORF411; ORF443; ORF548; ORF569; ORF601; ORF651; ORF654; ORF658; ORF659; ORF664; ORF665; ORF694; ORF698; ORF704; ORF760; ORF762; ORF763; ORF786; ORF787; ORF788; ORF801; ORF802; ORF812; ORF819; ORF822; ORF870; ORF897; ORF898; ORF902; ORF908; ORF916; ORF954; ORF955; ORF961; ORF983; ORF996; ORF1007; ORF1012; ORF1013; ORF1014; ORF1015; ORF1038; ORF1137 and one of their representative fragments.

Preferably, the invention also relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* polypeptide or one of its representative fragments which is involved in the metabolism of amino acids or polypeptides, such as for example serine hydroxymethyl transferase or the proteins which load amino acids onto transfer RNAs, and in that they comprise a nucleotide sequence chosen from the following sequences:

ORF99; ORF111; ORF127; ORF134; ORF140; ORF174; ORF175; ORF176; ORF353; ORF377; ORF404; ORF523; ORF539; ORF559; ORF561; ORF586; ORF598; ORF609; ORF636; ORF687; ORF700; ORF701; ORF759; ORF790; ORF857; ORF861; ORF904; ORF936; ORF952; ORF962; ORF963; ORF964; ORF965; ORF991; ORF1003; ORF1004; ORF1005; ORF1018; ORF1067; ORF1110; ORF1111; ORF1112; ORF1114; ORF1121; ORF1122; ORF1123; ORF1124; ORF1125 and one of their representative fragments.

Preferably, the invention also relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* polypeptide or one of its representative fragments which is involved in the metabolism of polypeptides, such as for example protein kinases or proteases, and in that they comprise a nucleotide sequence chosen from the following sequences:

ORF4; ORF44; ORF45; ORF48; ORF54; ORF112; ORF130; ORF155; ORF163; ORF212; ORF257; ORF307; ORF343; ORF405; ORF416; ORF458; ORF540; ORF541; ORF542; ORF543; ORF544; ORF560; ORF594; ORF652; ORF699; ORF723; ORF747; ORF817; ORF827; ORF871; ORF909; ORF910; ORF911; ORF912; ORF1023; ORF1051; ORF1052; ORF1081 and one of their representative fragments.

Preferably, the invention also relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* polypeptide or one of its representative fragments which is involved in the metabolism of fatty acids, such as for example succinyl-CoA-synthesizing proteins or phosphatidylserine synthetase, and in that they comprise a nucleotide sequence chosen from the following sequences:

ORF76; ORF284; ORF308; ORF309; ORF310; ORF311; ORF312; ORF425; ORF433; ORF565; ORF688; ORF690; ORF691; ORF767; ORF797; ORF894; ORF895; ORF994; ORF1020; ORF1030; ORF1033; ORF1034; ORF1046; ORF1047; ORF1057 and one of their representative fragments.

Preferably, the invention also relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* polypeptide or one of its

representative fragments which is involved in the synthesis of the wall, such as for example KDO transferase, and the proteins responsible for the attachment of certain sugars onto the exposed proteins, and in that they comprise a nucleotide sequence chosen from the following sequences: ORF49; ORF50; ORF177; ORF178; ORF245; ORF610; ORF972; ORF974; ORF978; ORF1037 and one of their representative fragments.

Preferably, the invention also relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* polypeptide or one of its representative fragments which is involved in the transcription, translation and/or maturation process, such as for example initiation factors, RNA polymerases or certain chaperone proteins, and in that they comprise a nucleotide sequence chosen from the following sequences: ORF90; ORF92; ORF131; ORF151; ORF199; ORF333; ORF334; ORF336; ORF379; ORF589; ORF590; ORF619; ORF630; ORF649; ORF739; ORF741; ORF806; ORF821; ORF843; ORF968; ORF971; ORF1061 and one of their representative fragments.

Preferably, the invention also relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* ribosomal polypeptide or one of its representative fragments, such as for example the ribosomal proteins L21, L27 and S10, and in that they comprise a nucleotide sequence chosen from the following sequences: ORF93; ORF94; ORF95; ORF136; ORF259; ORF332; ORF348; ORF583; ORF584; ORF588; ORF591; ORF592; ORF663; ORF666; ORF667; ORF669; ORF670; ORF671; ORF672; ORF673; ORF674; ORF675; ORF676; ORF677; ORF678; ORF679; ORF680; ORF681; ORF683; ORF684; ORF738; ORF781; ORF1008; ORF1024; ORF1025; ORF1066 and one of their representative fragments.

Preferably, the invention also relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* transport polypeptide or one of its representative fragments, such as for example the proteins for transporting amino acids, sugars and certain oligopeptides, and in that they comprise a nucleotide sequence chosen from the following sequences: ORF40; ORF41; ORF52; ORF105; ORF106; ORF107; ORF109; ORF133; ORF210; ORF211; ORF214; ORF215; ORF216; ORF217; ORF218; ORF219; ORF220; ORF223; ORF242; ORF260; ORF293; ORF299; ORF366; ORF369; ORF575; ORF602; ORF638; ORF639; ORF640; ORF643; ORF653; ORF702; ORF703; ORF724; ORF732; ORF855; ORF856; ORF901; ORF906; ORF933; ORF942; ORF1043; ORF1086; ORF1105 and one of their representative fragments.

Preferably, the invention also relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* polypeptide or one of its representative fragments which is involved in the virulence process, such as for example the proteins analogous to the *Escherichia coli* vacB protein, and in that they comprise a nucleotide sequence chosen from the following sequences:

ORF546; ORF550; ORF778; ORF779; ORF886 and one of their representative fragments.

Preferably, the invention also relates to the nucleotide sequences according to the invention, characterized in that they encode a *Chlamydia pneumoniae* polypeptide or one of its representative fragments which is involved in the secretory system and/or which is secreted, such as for example proteins homologous to proteins in the secretory system of certain bacteria such as the Salmonellae or the Yersiniae, and in that they comprise a nucleotide sequence chosen from the following sequences:

ORF751; ORF874; ORF875; ORF876; ORF883; ORF884; ORF885 and one of their representative fragments.

Preferably, the invention also relates to a nucleotide sequence according to the invention, characterized in that they encode a polypeptide specific to *Chlamydia pneumoniae* or one of its representative fragments (with a Blast E value of $>10^{-5}$), and in that they comprise a nucleotide sequence chosen from the following sequences:

ORF7; ORF8; ORF17; ORF18; ORF19; ORF20; ORF22; ORF23; ORF24; ORF51; ORF60; ORF63; ORF65; ORF66; ORF67; ORF83; ORF84; ORF86; ORF87; ORF125; ORF143; ORF144; ORF179; ORF182; ORF184; ORF185; ORF187; ORF221; ORF252; ORF254;; ORF278; ORF279; ORF387; ORF388; ORF397; ORF1048; ORF1049; ORF1050; ORF1128; ORF1130; ORF1131 and one of their representative fragments.

Also forming part of the invention are polypeptides encoded by the polynucleotides of the invention, as well as fusion polypeptides comprising such polypeptides. In one embodiment, the polypeptides and fusion polypeptides immunoreact with seropositive serum of an individual infected with *Chlamydia pneumoniae*. For example, described below, are polypeptide sequences exhibiting particularly preferable characteristics. For each group of preferred polypeptides described below, it is to be understood that in addition to the individual polypeptides listed, in instances wherein such polypeptides are encoded as part of "combined" ORFs, such "combined" polypeptides are also to be included within the preferred group.

The subject of the invention is also a polypeptide according to the invention, characterized in that it is a polypeptide of the cellular envelope, preferably of the outer cellular envelope, of *Chlamydia pneumoniae* or one of its representative fragments. According to the invention, the said polypeptide is preferably chosen from the polypeptides having the following sequences:

SEQ ID No. 15; SEQ ID No. 25; SEQ ID No. 26; SEQ ID No. 27; SEQ ID No. 28; SEQ ID No. 29; SEQ ID No. 30; SEQ ID No. 31; SEQ ID No. 32; SEQ ID No. 33; SEQ ID No. 35; SEQ ID No. 68; SEQ ID No. 124; SEQ ID No. 275; SEQ ID No. 291; SEQ ID No. 294; SEQ ID No. 327; SEQ ID No. 342; SEQ ID No. 364; SEQ ID No. 374; SEQ ID No. 380; SEQ ID No. 414; SEQ ID No. 439; SEQ ID No. 466; SEQ ID No. 467; SEQ ID No. 468; SEQ ID No. 469; SEQ ID No. 470; SEQ ID No. 472; SEQ ID No. 474; SEQ ID No. 476; SEQ ID No. 477; SEQ ID No. 478; SEQ ID No. 479;

SEQ ID No. 480; SEQ ID No. 482; SEQ ID No. 485; SEQ ID No. 500; SEQ ID No. 501; SEQ ID No. 503; SEQ ID No. 504; SEQ ID No. 505; SEQ ID No. 506; SEQ ID No. 520; SEQ ID No. 578; SEQ ID No. 580; SEQ ID No. 581; SEQ ID No. 595; SEQ ID No. 596; SEQ ID No. 597; SEQ ID No. 737; SEQ ID No. 830; SEQ ID No. 834; SEQ ID No. 836; SEQ ID No. 893; SEQ ID No. 917; SEQ ID No. 932; SEQ ID No. 976; SEQ ID No. 1035; SEQ ID No. 1045; SEQ ID No. 1090 and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, characterized in that it is a *Chlamydia pneumoniae* transmembrane polypeptide or one of its representative fragments, having between 1 and 3 transmembrane domains, and in that it is chosen from the polypeptides having the following sequences:

SEQ ID No. 2; SEQ ID No. 3; SEQ ID No. 6; SEQ ID No. 9; SEQ ID No. 10; SEQ ID No. 11; SEQ ID No. 13; SEQ ID No. 14; SEQ ID No. 16; SEQ ID No. 18; SEQ ID No. 19; SEQ ID No. 20; SEQ ID No. 21; SEQ ID No. 22; SEQ ID No. 25; SEQ ID No. 27; SEQ ID No. 28; SEQ ID No. 29; SEQ ID No. 30; SEQ ID No. 31; SEQ ID No. 32; SEQ ID No. 33; SEQ ID No. 34; SEQ ID No. 35; SEQ ID No. 37; SEQ ID No. 39; SEQ ID No. 41; SEQ ID No. 42; SEQ ID No. 44; SEQ ID No. 45; SEQ ID No. 46; SEQ ID No. 47; SEQ ID No. 48; SEQ ID No. 49; SEQ ID No. 50; SEQ ID No. 53; SEQ ID No. 54; SEQ ID No. 56; SEQ ID No. 57; SEQ ID No. 59; SEQ ID No. 60; SEQ ID No. 61; SEQ ID No. 62; SEQ ID No. 63; SEQ ID No. 64; SEQ ID No. 65; SEQ ID No. 66; SEQ ID No. 69;; SEQ ID No. 72; SEQ ID No. 73; SEQ ID No. 74; SEQ ID No. 76; SEQ ID No. 77; SEQ ID No. 78; SEQ ID No. 79; SEQ ID No. 80; SEQ ID No. 82; SEQ ID No. 84; SEQ ID No. 85; SEQ ID No. 86; SEQ ID No. 88; SEQ ID No. 89; SEQ ID No. 90; SEQ ID No. 91; SEQ ID No. 92; SEQ ID No. 93; SEQ ID No. 95; SEQ ID No. 96; SEQ ID No. 98; SEQ ID No. 99; SEQ ID No. 100; SEQ ID No. 101; SEQ ID No. 102; SEQ ID No. 103; SEQ ID No. 104; SEQ ID No. 105; SEQ ID No. 106; SEQ ID No. 107; SEQ ID No. 108; SEQ ID No. 114; SEQ ID No. 117; SEQ ID No. 118; SEQ ID No. 122; SEQ ID No. 123; SEQ ID No. 124; SEQ ID No. 125; SEQ ID No. 129; SEQ ID No. 130; SEQ ID No. 131; SEQ ID No. 132; SEQ ID No. 133; SEQ ID No. 134; SEQ ID No. 135; SEQ ID No. 137; SEQ ID No. 138; SEQ ID No. 139; SEQ ID No. 140; SEQ ID No. 141; SEQ ID No. 142; SEQ ID No. 143; SEQ ID No. 145; SEQ ID No. 146; SEQ ID No. 147; SEQ ID No. 150; SEQ ID No. 151; SEQ ID No. 152; SEQ ID No. 156; SEQ ID No. 157; SEQ ID No. 158; SEQ ID No. 159; SEQ ID No. 160; SEQ ID No. 161; SEQ ID No. 162; SEQ ID No. 164; SEQ ID No. 166; SEQ ID No. 167; SEQ ID No. 170; SEQ ID No. 173; SEQ ID No. 175; SEQ ID No. 176; SEQ ID No. 178; SEQ ID No. 179; SEQ ID No. 180; SEQ ID No. 182; SEQ ID No. 183; SEQ ID No. 184; SEQ ID No. 185; SEQ ID No. 186; SEQ ID No. 187; SEQ ID No. 188; SEQ ID No. 189; SEQ ID No. 190; SEQ ID No. 191; SEQ ID No. 192; SEQ ID No. 194; SEQ ID No. 195; SEQ ID No. 196; SEQ ID No. 197; SEQ ID No. 198; SEQ ID No. 199; SEQ ID No. 200; SEQ ID No. 201; SEQ ID No. 202; SEQ ID No. 205; SEQ ID No. 207; SEQ ID No. 208; SEQ ID No. 209; SEQ ID No. 210; SEQ ID No. 212; SEQ ID

No. 215; SEQ ID No. 219; SEQ ID No. 220; SEQ ID No. 224; SEQ ID No. 226; SEQ ID No. 227; SEQ ID No. 228; SEQ ID No. 231; SEQ ID No. 232; SEQ ID No. 233; SEQ ID No. 234; SEQ ID No. 235; SEQ ID No. 236; SEQ ID No. 238; SEQ ID No. 239; SEQ ID No. 240; SEQ ID No. 241; SEQ ID No. 242; SEQ ID No. 244; SEQ ID No. 247; SEQ ID No. 251; SEQ ID No. 252; SEQ ID No. 253; SEQ ID No. 255; SEQ ID No. 256; SEQ ID No. 257; SEQ ID No. 258; SEQ ID No. 260; SEQ ID No. 262; SEQ ID No. 263; SEQ ID No. 266; SEQ ID No. 267; SEQ ID No. 268; SEQ ID No. 269; SEQ ID No. 270; SEQ ID No. 273; SEQ ID No. 274; SEQ ID No. 276; SEQ ID No. 278; SEQ ID No. 279; SEQ ID No. 280; SEQ ID No. 281; SEQ ID No. 282; SEQ ID No. 283; SEQ ID No. 284; SEQ ID No. 286; SEQ ID No. 287; SEQ ID No. 289; SEQ ID No. 290; SEQ ID No. 291; SEQ ID No. 293; SEQ ID No. 294; SEQ ID No. 297; SEQ ID No. 304; SEQ ID No. 305; SEQ ID No. 307; SEQ ID No. 308; SEQ ID No. 309; SEQ ID No. 310; SEQ ID No. 311; SEQ ID No. 313; SEQ ID No. 314; SEQ ID No. 315; SEQ ID No. 316; SEQ ID No. 318; SEQ ID No. 319; SEQ ID No. 320; SEQ ID No. 321; SEQ ID No. 322; SEQ ID No. 323; SEQ ID No. 324; SEQ ID No. 325; SEQ ID No. 326; SEQ ID No. 331; SEQ ID No. 332; SEQ ID No. 336; SEQ ID No. 338; SEQ ID No. 339; SEQ ID No. 341; SEQ ID No. 344; SEQ ID No. 345; SEQ ID No. 346; SEQ ID No. 350; SEQ ID No. 352; SEQ ID No. 353; SEQ ID No. 356; SEQ ID No. 357; SEQ ID No. 358; SEQ ID No. 359; SEQ ID No. 360; SEQ ID No. 362; SEQ ID No. 365; SEQ ID No. 366; SEQ ID No. 367; SEQ ID No. 370; SEQ ID No. 372; SEQ ID No. 373; SEQ ID No. 376; SEQ ID No. 377; SEQ ID No. 378; SEQ ID No. 379; SEQ ID No. 381; SEQ ID No. 382; SEQ ID No. 383; SEQ ID No. 384; SEQ ID No. 385; SEQ ID No. 386; SEQ ID No. 387; SEQ ID No. 390; SEQ ID No. 392; SEQ ID No. 393; SEQ ID No. 394; SEQ ID No. 396; SEQ ID No. 398; SEQ ID No. 399; SEQ ID No. 400; SEQ ID No. 404; SEQ ID No. 408; SEQ ID No. 410; SEQ ID No. 411; SEQ ID No. 413; SEQ ID No. 416; SEQ ID No. 417; SEQ ID No. 418; SEQ ID No. 420; SEQ ID No. 422; SEQ ID No. 424; SEQ ID No. 427; SEQ ID No. 428; SEQ ID No. 429; SEQ ID No. 430; SEQ ID No. 431; SEQ ID No. 433; SEQ ID No. 434; SEQ ID No. 437; SEQ ID No. 440; SEQ ID No. 441; SEQ ID No. 442; SEQ ID No. 443; SEQ ID No. 444; SEQ ID No. 445; SEQ ID No. 447; SEQ ID No. 450; SEQ ID No. 451; SEQ ID No. 452; SEQ ID No. 455; SEQ ID No. 456; SEQ ID No. 459; SEQ ID No. 460; SEQ ID No. 461; SEQ ID No. 462; SEQ ID No. 463; SEQ ID No. 464; SEQ ID No. 465; SEQ ID No. 467; SEQ ID No. 469; SEQ ID No. 471; SEQ ID No. 474; SEQ ID No. 475; SEQ ID No. 476; SEQ ID No. 477; SEQ ID No. 479; SEQ ID No. 482; SEQ ID No. 483; SEQ ID No. 484; SEQ ID No. 485; SEQ ID No. 486; SEQ ID No. 487; SEQ ID No. 488; SEQ ID No. 491; SEQ ID No. 493; SEQ ID No. 494; SEQ ID No. 497; SEQ ID No. 498; SEQ ID No. 499; SEQ ID No. 503; SEQ ID No. 508; SEQ ID No. 509; SEQ ID No. 510; SEQ ID No. 512; SEQ ID No. 514; SEQ ID No. 515; SEQ ID No. 516; SEQ ID No. 517; SEQ ID No. 518; SEQ ID No. 520; SEQ ID No. 521; SEQ ID No. 523; SEQ ID No. 525; SEQ ID No. 527; SEQ ID No. 528; SEQ ID No. 529; SEQ ID No. 530; SEQ ID No. 531; SEQ ID No. 533; SEQ ID No. 534; SEQ ID No. 535; SEQ ID No. 536; SEQ ID No. 537; SEQ ID No. 540; SEQ ID No. 541; SEQ ID No. 543; SEQ ID No. 544; SEQ ID

No. 545; SEQ ID No. 546; SEQ ID No. 548; SEQ ID No. 549; SEQ ID No. 551; SEQ ID No. 553; SEQ ID No. 554; SEQ ID No. 555; SEQ ID No. 556; SEQ ID No. 557; SEQ ID No. 558; SEQ ID No. 559; SEQ ID No. 560; SEQ ID No. 562; SEQ ID No. 563; SEQ ID No. 564; SEQ ID No. 565; SEQ ID No. 566; SEQ ID No. 569; SEQ ID No. 571; SEQ ID No. 573; SEQ ID No. 576; SEQ ID No. 577; SEQ ID No. 581; SEQ ID No. 583; SEQ ID No. 584; SEQ ID No. 585; SEQ ID No. 586; SEQ ID No. 588; SEQ ID No. 591; SEQ ID No. 592; SEQ ID No. 594; SEQ ID No. 595; SEQ ID No. 596; SEQ ID No. 597; SEQ ID No. 599; SEQ ID No. 600; SEQ ID No. 603; SEQ ID No. 605; SEQ ID No. 608; SEQ ID No. 614; SEQ ID No. 615; SEQ ID No. 620; SEQ ID No. 621; SEQ ID No. 622; SEQ ID No. 623; SEQ ID No. 624; SEQ ID No. 625; SEQ ID No. 629; SEQ ID No. 630; SEQ ID No. 631; SEQ ID No. 633; SEQ ID No. 634; SEQ ID No. 637; SEQ ID No. 642; SEQ ID No. 644; SEQ ID No. 645; SEQ ID No. 647; SEQ ID No. 648; SEQ ID No. 652; SEQ ID No. 654; SEQ ID No. 655; SEQ ID No. 657; SEQ ID No. 658; SEQ ID No. 659; SEQ ID No. 660; SEQ ID No. 661; SEQ ID No. 664; SEQ ID No. 665; SEQ ID No. 666; SEQ ID No. 667; SEQ ID No. 670; SEQ ID No. 671; SEQ ID No. 672; SEQ ID No. 673; SEQ ID No. 674; SEQ ID No. 676; SEQ ID No. 679; SEQ ID No. 681; SEQ ID No. 684; SEQ ID No. 687; SEQ ID No. 688; SEQ ID No. 689; SEQ ID No. 690; SEQ ID No. 693; SEQ ID No. 694; SEQ ID No. 695; SEQ ID No. 696; SEQ ID No. 697; SEQ ID No. 698; SEQ ID No. 699; SEQ ID No. 700; SEQ ID No. 701; SEQ ID No. 703; SEQ ID No. 705; SEQ ID No. 706; SEQ ID No. 707; SEQ ID No. 708; SEQ ID No. 710; SEQ ID No. 712; SEQ ID No. 715; SEQ ID No. 716; SEQ ID No. 717; SEQ ID No. 718; SEQ ID No. 719; SEQ ID No. 721; SEQ ID No. 722; SEQ ID No. 723; SEQ ID No. 725; SEQ ID No. 726; SEQ ID No. 727; SEQ ID No. 728; SEQ ID No. 729; SEQ ID No. 730; SEQ ID No. 731; SEQ ID No. 733; SEQ ID No. 736; SEQ ID No. 737; SEQ ID No. 738; SEQ ID No. 740; SEQ ID No. 741; SEQ ID No. 742; SEQ ID No. 743; SEQ ID No. 747; SEQ ID No. 748; SEQ ID No. 750; SEQ ID No. 752; SEQ ID No. 754; SEQ ID No. 755; SEQ ID No. 756; SEQ ID No. 757; SEQ ID No. 759; SEQ ID No. 760; SEQ ID No. 761; SEQ ID No. 762; SEQ ID No. 763; SEQ ID No. 764; SEQ ID No. 765; SEQ ID No. 766; SEQ ID No. 767; SEQ ID No. 768; SEQ ID No. 772; SEQ ID No. 774; SEQ ID No. 775; SEQ ID No. 777; SEQ ID No. 781; SEQ ID No. 783; SEQ ID No. 788; SEQ ID No. 791; SEQ ID No. 792; SEQ ID No. 793; SEQ ID No. 794; SEQ ID No. 795; SEQ ID No. 796; SEQ ID No. 797; SEQ ID No. 798; SEQ ID No. 799; SEQ ID No. 802; SEQ ID No. 803; SEQ ID No. 806; SEQ ID No. 807; SEQ ID No. 808; SEQ ID No. 809; SEQ ID No. 810; SEQ ID No. 811; SEQ ID No. 813; SEQ ID No. 814; SEQ ID No. 815; SEQ ID No. 816; SEQ ID No. 817; SEQ ID No. 819; SEQ ID No. 820; SEQ ID No. 821; SEQ ID No. 823; SEQ ID No. 824; SEQ ID No. 827; SEQ ID No. 829; SEQ ID No. 830; SEQ ID No. 831; SEQ ID No. 833; SEQ ID No. 834; SEQ ID No. 835; SEQ ID No. 837; SEQ ID No. 844; SEQ ID No. 845; SEQ ID No. 846; SEQ ID No. 847; SEQ ID No. 848; SEQ ID No. 849; SEQ ID No. 850; SEQ ID No. 851; SEQ ID No. 852; SEQ ID No. 854; SEQ ID No. 855; SEQ ID No. 856; SEQ ID No. 857; SEQ ID No. 859; SEQ ID No. 860; SEQ ID No. 862; SEQ ID No. 865; SEQ ID No. 866; SEQ ID No. 868; SEQ ID No. 869; SEQ ID

No. 870; SEQ ID No. 871; SEQ ID No. 872; SEQ ID No. 874; SEQ ID No. 877; SEQ ID No. 878; SEQ ID No. 879; SEQ ID No. 880; SEQ ID No. 881; SEQ ID No. 882; SEQ ID No. 884; SEQ ID No. 885; SEQ ID No. 888; SEQ ID No. 889; SEQ ID No. 890; SEQ ID No. 891; SEQ ID No. 892; SEQ ID No. 894; SEQ ID No. 895; SEQ ID No. 896; SEQ ID No. 897; SEQ ID No. 899; SEQ ID No. 900; SEQ ID No. 902; SEQ ID No. 903; SEQ ID No. 904; SEQ ID No. 905; SEQ ID No. 909; SEQ ID No. 910; SEQ ID No. 912; SEQ ID No. 913; SEQ ID No. 914; SEQ ID No. 915; SEQ ID No. 917; SEQ ID No. 918; SEQ ID No. 919; SEQ ID No. 921; SEQ ID No. 923; SEQ ID No. 924; SEQ ID No. 926; SEQ ID No. 927; SEQ ID No. 928; SEQ ID No. 929; SEQ ID No. 930; SEQ ID No. 931; SEQ ID No. 937; SEQ ID No. 938; SEQ ID No. 939; SEQ ID No. 941; SEQ ID No. 943; SEQ ID No. 948; SEQ ID No. 951; SEQ ID No. 952; SEQ ID No. 953; SEQ ID No. 958; SEQ ID No. 960; SEQ ID No. 963; SEQ ID No. 964; SEQ ID No. 965; SEQ ID No. 968; SEQ ID No. 970; SEQ ID No. 974; SEQ ID No. 975; SEQ ID No. 977; SEQ ID No. 979; SEQ ID No. 980; SEQ ID No. 981; SEQ ID No. 983; SEQ ID No. 984; SEQ ID No. 985; SEQ ID No. 987; SEQ ID No. 989; SEQ ID No. 992; SEQ ID No. 993; SEQ ID No. 997; SEQ ID No. 998; SEQ ID No. 999; SEQ ID No. 1001; SEQ ID No. 1002; SEQ ID No. 1004; SEQ ID No. 1005; SEQ ID No. 1009; SEQ ID No. 1013; SEQ ID No. 1014; SEQ ID No. 1015; SEQ ID No. 1016; SEQ ID No. 1019; SEQ ID No. 1021; SEQ ID No. 1023; SEQ ID No. 1024; SEQ ID No. 1029; SEQ ID No. 1031; SEQ ID No. 1033; SEQ ID No. 1034; SEQ ID No. 1039; SEQ ID No. 1041; SEQ ID No. 1042; SEQ ID No. 1045; SEQ ID No. 1047; SEQ ID No. 1049; SEQ ID No. 1051; SEQ ID No. 1052; SEQ ID No. 1053; SEQ ID No. 1054; SEQ ID No. 1056; SEQ ID No. 1059; SEQ ID No. 1061; SEQ ID No. 1062; SEQ ID No. 1063; SEQ ID No. 1064; SEQ ID No. 1065; SEQ ID No. 1067; SEQ ID No. 1075; SEQ ID No. 1077; SEQ ID No. 1078; SEQ ID No. 1079; SEQ ID No. 1080; SEQ ID No. 1081; SEQ ID No. 1089; SEQ ID No. 1095; SEQ ID No. 1097; SEQ ID No. 1098; SEQ ID No. 1099; SEQ ID No. 1101; SEQ ID No. 1102; SEQ ID No. 1103; SEQ ID No. 1106; SEQ ID No. 1107; SEQ ID No. 1108; SEQ ID No. 1109; SEQ ID No. 1110; SEQ ID No. 1113; SEQ ID No. 1116; SEQ ID No. 1118; SEQ ID No. 1119; SEQ ID No. 1121; SEQ ID No. 1123; SEQ ID No. 1124; SEQ ID No. 1126; SEQ ID No. 1128; SEQ ID No. 1130; SEQ ID No. 1131; SEQ ID No. 1133; SEQ ID No. 1134; SEQ ID No. 1136; SEQ ID No. 1137 and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, characterized in that it is a *Chlamydia pneumoniae* transmembrane polypeptide or one of its respective fragments, having between 4 and 6 transmembrane domains, and in that it is chosen from the polypeptides having the following sequences:

SEQ ID No. 5; SEQ ID No. 7; SEQ ID No. 8; SEQ ID No. 15; SEQ ID No. 36; SEQ ID No. 38; SEQ ID No. 51; SEQ ID No. 55; SEQ ID No. 58; SEQ ID No. 67; SEQ ID No. 70; SEQ ID No. 81; SEQ ID No. 97; SEQ ID No. 110; SEQ ID No. 111; SEQ ID No. 115; SEQ ID No. 119; SEQ ID No. 126; SEQ ID No. 128; SEQ ID No. 148; SEQ ID No. 155; SEQ ID No. 163; SEQ ID

No. 165; SEQ ID No. 168; SEQ ID No. 169; SEQ ID No. 171; SEQ ID No. 172; SEQ ID No. 174; SEQ ID No. 177; SEQ ID No. 181; SEQ ID No. 193; SEQ ID No. 203; SEQ ID No. 213; SEQ ID No. 214; SEQ ID No. 216; SEQ ID No. 217; SEQ ID No. 221; SEQ ID No. 222; SEQ ID No. 225; SEQ ID No. 229; SEQ ID No. 243; SEQ ID No. 246; SEQ ID No. 248; SEQ ID No. 254; SEQ ID No. 261; SEQ ID No. 285; SEQ ID No. 288; SEQ ID No. 292; SEQ ID No. 296; SEQ ID No. 298; SEQ ID No. 299; SEQ ID No. 301; SEQ ID No. 303; SEQ ID No. 317; SEQ ID No. 328; SEQ ID No. 329; SEQ ID No. 351; SEQ ID No. 354; SEQ ID No. 355; SEQ ID No. 364; SEQ ID No. 371; SEQ ID No. 374; SEQ ID No. 375; SEQ ID No. 391; SEQ ID No. 395; SEQ ID No. 401; SEQ ID No. 403; SEQ ID No. 405; SEQ ID No. 409; SEQ ID No. 414; SEQ ID No. 419; SEQ ID No. 421; SEQ ID No. 423; SEQ ID No. 425; SEQ ID No. 438; SEQ ID No. 448; SEQ ID No. 453; SEQ ID No. 458; SEQ ID No. 466; SEQ ID No. 468; SEQ ID No. 470; SEQ ID No. 480; SEQ ID No. 489; SEQ ID No. 490; SEQ ID No. 496; SEQ ID No. 501; SEQ ID No. 504; SEQ ID No. 505; SEQ ID No. 506; SEQ ID No. 511; SEQ ID No. 513; SEQ ID No. 519; SEQ ID No. 526; SEQ ID No. 532; SEQ ID No. 538; SEQ ID No. 539; SEQ ID No. 547; SEQ ID No. 550; SEQ ID No. 561; SEQ ID No. 568; SEQ ID No. 570; SEQ ID No. 574; SEQ ID No. 578; SEQ ID No. 579; SEQ ID No. 580; SEQ ID No. 582; SEQ ID No. 589; SEQ ID No. 593; SEQ ID No. 598; SEQ ID No. 601; SEQ ID No. 604; SEQ ID No. 610; SEQ ID No. 613; SEQ ID No. 617; SEQ ID No. 626; SEQ ID No. 632; SEQ ID No. 635; SEQ ID No. 638; SEQ ID No. 640; SEQ ID No. 641; SEQ ID No. 646; SEQ ID No. 649; SEQ ID No. 650; SEQ ID No. 651; SEQ ID No. 686; SEQ ID No. 711; SEQ ID No. 724; SEQ ID No. 732; SEQ ID No. 734; SEQ ID No. 744; SEQ ID No. 745; SEQ ID No. 749; SEQ ID No. 751; SEQ ID No. 769; SEQ ID No. 770; SEQ ID No. 771; SEQ ID No. 773; SEQ ID No. 776; SEQ ID No. 779; SEQ ID No. 780; SEQ ID No. 785; SEQ ID No. 787; SEQ ID No. 789; SEQ ID No. 801; SEQ ID No. 805; SEQ ID No. 812; SEQ ID No. 822; SEQ ID No. 825; SEQ ID No. 826; SEQ ID No. 839; SEQ ID No. 841; SEQ ID No. 843; SEQ ID No. 853; SEQ ID No. 861; SEQ ID No. 875; SEQ ID No. 876; SEQ ID No. 886; SEQ ID No. 893; SEQ ID No. 898; SEQ ID No. 906; SEQ ID No. 907; SEQ ID No. 908; SEQ ID No. 920; SEQ ID No. 922; SEQ ID No. 925; SEQ ID No. 933; SEQ ID No. 935; SEQ ID No. 936; SEQ ID No. 944; SEQ ID No. 946; SEQ ID No. 947; SEQ ID No. 954; SEQ ID No. 959; SEQ ID No. 961; SEQ ID No. 966; SEQ ID No. 967; SEQ ID No. 972; SEQ ID No. 978; SEQ ID No. 995; SEQ ID No. 996; SEQ ID No. 1000; SEQ ID No. 1003; SEQ ID No. 1010; SEQ ID No. 1011; SEQ ID No. 1012; SEQ ID No. 1017; SEQ ID No. 1020; SEQ ID No. 1030; SEQ ID No. 1036; SEQ ID No. 1038; SEQ ID No. 1043; SEQ ID No. 1046; SEQ ID No. 1048; SEQ ID No. 1050; SEQ ID No. 1058; SEQ ID No. 1071; SEQ ID No. 1073; SEQ ID No. 1084; SEQ ID No. 1085; SEQ ID No. 1086; SEQ ID No. 1087; SEQ ID No. 1091; SEQ ID No. 1092; SEQ ID No. 1094; SEQ ID No. 1096; SEQ ID No. 1100; SEQ ID No. 1104; SEQ ID No. 1111; SEQ ID No. 1112; SEQ ID No. 1114; SEQ ID No. 1117; SEQ ID No. 1122; SEQ ID No. 1125 and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention,

characterized in that it is a *Chlamydia pneumoniae* transmembrane polypeptide or one of its representative fragments, having at least 7 transmembrane domains, and in that it is chosen from the polypeptides having the following sequences:

SEQ ID No. 17; SEQ ID No. 52; SEQ ID No. 68; SEQ ID No. 83; SEQ ID No. 87; SEQ ID No. 109; SEQ ID No. 112; SEQ ID No. 113; SEQ ID No. 120; SEQ ID No. 121; SEQ ID No. 127; SEQ ID No. 153; SEQ ID No. 204; SEQ ID No. 211; SEQ ID No. 218; SEQ ID No. 223; SEQ ID No. 275; SEQ ID No. 277; SEQ ID No. 295; SEQ ID No. 300; SEQ ID No. 302; SEQ ID No. 306; SEQ ID No. 327; SEQ ID No. 335; SEQ ID No. 342; SEQ ID No. 343; SEQ ID No. 347; SEQ ID No. 349; SEQ ID No. 361; SEQ ID No. 363; SEQ ID No. 369; SEQ ID No. 380; SEQ ID No. 388; SEQ ID No. 389; SEQ ID No. 397; SEQ ID No. 415; SEQ ID No. 432; SEQ ID No. 439; SEQ ID No. 446; SEQ ID No. 449; SEQ ID No. 472; SEQ ID No. 478; SEQ ID No. 500; SEQ ID No. 522; SEQ ID No. 524; SEQ ID No. 567; SEQ ID No. 575; SEQ ID No. 602; SEQ ID No. 606; SEQ ID No. 609; SEQ ID No. 636; SEQ ID No. 639; SEQ ID No. 643; SEQ ID No. 653; SEQ ID No. 668; SEQ ID No. 692; SEQ ID No. 702; SEQ ID No. 704; SEQ ID No. 713; SEQ ID No. 720; SEQ ID No. 778; SEQ ID No. 784; SEQ ID No. 800; SEQ ID No. 836; SEQ ID No. 838; SEQ ID No. 842; SEQ ID No. 864; SEQ ID No. 867; SEQ ID No. 883; SEQ ID No. 901; SEQ ID No. 916; SEQ ID No. 932; SEQ ID No. 934; SEQ ID No. 940; SEQ ID No. 942; SEQ ID No. 950; SEQ ID No. 956; SEQ ID No. 971; SEQ ID No. 973; SEQ ID No. 976; SEQ ID No. 988; SEQ ID No. 994; SEQ ID No. 1018; SEQ ID No. 1028; SEQ ID No. 1035; SEQ ID No. 1037; SEQ ID No. 1044; SEQ ID No. 1055; SEQ ID No. 1057; SEQ ID No. 1068; SEQ ID No. 1069; SEQ ID No. 1070; SEQ ID No. 1072; SEQ ID No. 1082; SEQ ID No. 1088; SEQ ID No. 1105; SEQ ID No. 1132; SEQ ID No. 1135 and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, in that it is a *Chlamydia pneumoniae* surface exposed polypeptide or one of its representative fragments, and in that it is chosen from the polypeptides having the following sequences:

SEQ ID No. 15, SEQ ID No. 25, SEQ ID No. 26, SEQ ID No. 27, SEQ ID No. 28, SEQ ID No. 29, SEQ ID No. 30, SEQ ID No. 31, SEQ ID No. 32, SEQ ID No. 33, SEQ ID No. 35, SEQ ID No. 36, SEQ ID No. 1257, SEQ ID No. 280, SEQ ID No. 291, SEQ ID No. 314, SEQ ID No. 354, SEQ ID No. 380, SEQ ID No. 1266, SEQ ID No. 466, SEQ ID No. 467, SEQ ID No. 468, SEQ ID No. 469, SEQ ID No. 470, SEQ ID No. 472, SEQ ID No. 474, SEQ ID No. 476, SEQ ID No. 477, SEQ ID No. 478, SEQ ID No. 479, SEQ ID No. 480, SEQ ID No. 482, SEQ ID No. 483, SEQ ID No. 485, SEQ ID No. 486, SEQ ID No. 500, SEQ ID No. 501, SEQ ID No. 503, SEQ ID No. 504, SEQ ID No. 505, SEQ ID No. 506, SEQ ID No. 507, SEQ ID No. 1268, SEQ ID No. 1269, SEQ ID No. 543, SEQ ID No. 544, SEQ ID No. 578, SEQ ID No. 579, SEQ ID No. 580, SEQ ID No. 581, SEQ ID No. 595, SEQ ID No. 596, SEQ ID No. 597, SEQ ID No. 1271, SEQ ID No. 633, SEQ ID No. 637, SEQ ID No. 699, SEQ ID No. 706, SEQ ID No. 737, SEQ ID No. 744, SEQ ID No. 1273, SEQ ID No. 751, SEQ ID No. 775, SEQ ID No. 776, SEQ ID No. 777, SEQ ID No. 793, SEQ ID No. 815, SEQ ID No.

830, SEQ ID No. 1221, SEQ ID No. 849, SEQ ID No. 851, SEQ ID No. 852, SEQ ID No. 874, SEQ ID No. 891, SEQ ID No. 922, SEQ ID No. 940, SEQ ID No. 1231, SEQ ID No. 1281, SEQ ID No. 1035, SEQ ID No. 1079, SEQ ID No. 1087, SEQ ID No. 1108, and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, characterized in that it is a *Chlamydia pneumoniae* lipoprotein or one of its representative fragments, and in that it is chosen from the polypeptides having the following sequences:

SEQ ID No. 3, SEQ ID No. 10, SEQ ID No. 11, SEQ ID No. 16, SEQ ID No. 1254, SEQ ID No. 1255, SEQ ID No. 38, SEQ ID No. 1256, SEQ ID No. 62, SEQ ID No. 85, SEQ ID No. 1258, SEQ ID No. 115, SEQ ID No. 1151, SEQ ID No. 151, SEQ ID No. 1259, SEQ ID No. 173, SEQ ID No. 1261, SEQ ID No. 186, SEQ ID No. 194, SEQ ID No. 205, SEQ ID No. 214, SEQ ID No. 216, SEQ ID No. 217, SEQ ID No. 238, SEQ ID No. 1177, SEQ ID No. 280, SEQ ID No. 291, SEQ ID No. 317, SEQ ID No. 327, SEQ ID No. 354, SEQ ID No. 364, SEQ ID No. 367, SEQ ID No. 414, SEQ ID No. 432, SEQ ID No. 1192, SEQ ID No. 460, SEQ ID No. 1267, SEQ ID No. 1268, SEQ ID No. 520, SEQ ID No. 536, SEQ ID No. 1270, SEQ ID No. 576, SEQ ID No. 597, SEQ ID No. 603, SEQ ID No. 609, SEQ ID No. 637, SEQ ID No. 1272, SEQ ID No. 652, SEQ ID No. 1213, SEQ ID No. 699, SEQ ID No. 705, SEQ ID No. 706, SEQ ID No. 708, SEQ ID No. 711, SEQ ID No. 727, SEQ ID No. 1274, SEQ ID No. 800, SEQ ID No. 814, SEQ ID No. 825, SEQ ID No. 829, SEQ ID No. 830, SEQ ID No. 831, SEQ ID No. 844, SEQ ID No. 849, SEQ ID No. 1275, SEQ ID No. 1276, SEQ ID No. 1277, SEQ ID No. 872, SEQ ID No. 878, SEQ ID No. 880, SEQ ID No. 891, SEQ ID No. 892, SEQ ID No. 1278, SEQ ID No. 1279, SEQ ID No. 1280, SEQ ID No. 941, SEQ ID No. 942, SEQ ID No. 1282, SEQ ID No. 1283, SEQ ID No. 952, SEQ ID No. 988, SEQ ID No. 998, SEQ ID No. 1009, SEQ ID No. 1285, SEQ ID No. 1235, SEQ ID No. 1028, SEQ ID No. 1056, SEQ ID No. 1070, SEQ ID No. 1287, SEQ ID No. 1087, SEQ ID No. 1288, SEQ ID No. 1289, SEQ ID No. 1098, SEQ ID No. 1246, SEQ ID No. 1291, SEQ ID No. 1108, SEQ ID No. 1109, SEQ ID No. 1112, SEQ ID No. 1133, and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, in that it is a *Chlamydia pneumoniae* polypeptide involved in lipopolysaccharide (LPS) biosynthesis, and in that it is chosen from the polypeptides having the following sequences:

SEQ ID No. 316, SEQ ID No. 564, SEQ ID No. 610, SEQ ID No. 647, SEQ ID No. 1211, SEQ ID No. 688, SEQ ID No. 924, and one of their representative fragments.

Preferably, the invention relates to additional LPS-related polypeptides according to the invention, in that it is:

(a) a *Chlamydia pneumoniae* KDO (3-deoxy-D-manno-octylosonic acid)-related polypeptide or one of its representative fragments, and in that it is chosen from the polypeptides having the following sequences: SEQ ID No. 177, SEQ ID No. 1156, SEQ ID No. 245, SEQ ID No. 767, and one of their representative fragments;

(b) a *Chlamydia pneumoniae* phosphomannomutase-related polypeptide or one of its representative fragments, and in that it is chosen from the polypeptides having the following sequences: SEQ ID No. 74, and its representative fragment;

(c) a *Chlamydia pneumoniae* phosphoglucomutase-related polypeptide or one of its representative fragments, and in that it is chosen from the polypeptides having the following sequences: SEQ ID No. 1286, SEQ ID No. 1039, and its representative fragment; and

(d) a *Chlamydia pneumoniae* lipid A component-related polypeptide or one of its representative fragments, and in that it is chosen from the polypeptides having the following sequences: SEQ ID No. 689, SEQ ID No. 690, SEQ ID No. 691, SEQ ID No. 1037, and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, in that it is a *Chlamydia pneumoniae* polypeptide or one of its representative fragments that contains an RGD sequence and is also an outer membrane protein, and in that it is chosen from the polypeptides having the following sequences: SEQ ID No. 468 and its representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, in that it is a *Chlamydia pneumoniae* polypeptide or one of its representative fragments that contains an RGD sequence that shows homology to cds1, cds2, and copN type III virulence loci in *Chlamydia Psitacci*, and in that it is chosen from the polypeptides having the following sequences:
SEQ ID No. 350 and its representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, in that it is a *Chlamydia pneumoniae* polypeptide or one of its representative fragments that is cysteine-rich and contains RGD sequence, and in that it is chosen from the polypeptides having the following sequences: SEQ ID No. 1290, SEQ ID No. 6846, SEQ ID No. 6848, and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, in that it is a *Chlamydia pneumoniae* outer membrane polypeptide that contains cysteines in their first 30 amino acids and also contain an RGD sequence, and in that it is chosen from the polypeptides having the following sequences:
SEQ ID No. 105, SEQ ID No. 106, SEQ ID No. 114, SEQ ID No. 170, SEQ ID No. 171, SEQ ID No. 1264, SEQ ID No. 268, SEQ ID No. 1265, SEQ ID No. 350, SEQ ID No. 393, SEQ ID No. 394, SEQ ID No. 451, SEQ ID No. 452, SEQ ID No. 453, SEQ ID No. 473, SEQ ID No. 499, SEQ ID No. 515, SEQ ID No. 519, SEQ ID No. 525, SEQ ID No. 526, SEQ ID No. 538, SEQ ID No. 611, SEQ ID No. 645, SEQ ID No. 686, SEQ ID No. 700, SEQ ID No. 746, SEQ ID No. 755, SEQ ID No. 756, SEQ ID No. 757, SEQ ID No. 789, SEQ ID No. 814, SEQ ID No. 855, SEQ ID No. 856, SEQ ID No. 878, SEQ ID No. 957, SEQ ID No. 958, SEQ ID No. 989, SEQ ID No. 1290, and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, in that it is a

Chlamydia pneumoniae polypeptide or one of its representative fragments that contains RGD sequences homologous to *Chlamydia trachomatis* polypeptides containing RGD sequences, and in that it is chosen from the polypeptides having the following sequences:
SEQ ID No. 114, SEQ ID No. 468, SEQ ID No. 755, SEQ ID No. 756, SEQ ID No. 757, SEQ ID No. 855, SEQ ID No. 856, SEQ ID No. 905, SEQ ID No. 913, SEQ ID No. 914, SEQ ID No. 915, and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, in that it is a *Chlamydia pneumoniae* Type III and non-Type III secreted polypeptide or one of its representative fragments, and in that it is chosen from the polypeptides having the following sequences:
SEQ ID No. 25, SEQ ID No. 28, SEQ ID No. 29, SEQ ID No. 33, SEQ ID No. 308, SEQ ID No. 309, SEQ ID No. 343, SEQ ID No. 344, SEQ ID No. 345, SEQ ID No. 367, SEQ ID No. 414, SEQ ID No. 415, SEQ ID No. 480, SEQ ID No. 550, SEQ ID No. 579, SEQ ID No. 580, SEQ ID No. 581, SEQ ID No. 597, SEQ ID No. 699, SEQ ID No. 744, SEQ ID No. 751, SEQ ID No. 776, SEQ ID No. 866, SEQ ID No. 874, SEQ ID No. 883, SEQ ID No. 884, SEQ ID No. 888, SEQ ID No. 891, SEQ ID No. 6845, and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, in that it is a *Chlamydia pneumoniae* cell wall anchored surface polypeptide or one of its representative fragments, and in that it is chosen from the polypeptides having the following sequences:
SEQ ID No. 267, SEQ ID No. 271, SEQ ID No. 419, SEQ ID No. 590, SEQ ID No. 932, SEQ ID No. 6844, SEQ ID No. 6847, and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, in that it is a *Chlamydia pneumoniae* polypeptide or one of its representative fragments not found in *Chlamydia trachomatis* (Blastp $P>e^{-10}$), and in that it is chosen from the polypeptides having the following sequences:
SEQ ID No. 7, SEQ ID No. 8, SEQ ID No. 9, SEQ ID No. 16, SEQ ID No. 17, SEQ ID No. 18, SEQ ID No. 19, SEQ ID No. 20, SEQ ID No. 21, SEQ ID No. 22, SEQ ID No. 1254, SEQ ID No. 23, SEQ ID No. 1255, SEQ ID No. 24, SEQ ID No. 1139, SEQ ID No. 1140, SEQ ID No. 46, SEQ ID No. 47, SEQ ID No. 51, SEQ ID No. 60, SEQ ID No. 1256, SEQ ID No. 61, SEQ ID No. 62, SEQ ID No. 63, SEQ ID No. 64, SEQ ID No. 1257, SEQ ID No. 65, SEQ ID No. 66, SEQ ID No. 67, SEQ ID No. 68, SEQ ID No. 1143, SEQ ID No. 1145, SEQ ID No. 83, SEQ ID No. 84, SEQ ID No. 1146, SEQ ID No. 85, SEQ ID No. 86, SEQ ID No. 87, SEQ ID No. 1258, SEQ ID No. 116, SEQ ID No. 117, SEQ ID No. 125, SEQ ID No. 1148, SEQ ID No. 143, SEQ ID No. 1150, SEQ ID No. 1151, SEQ ID No. 144, SEQ ID No. 145, SEQ ID No. 147, SEQ ID No. 148, SEQ ID No. 149, SEQ ID No. 150, SEQ ID No. 152, SEQ ID No. 1259, SEQ ID No. 162, SEQ ID No. 166, SEQ ID No. 1154, SEQ ID No. 167, SEQ ID No. 1261, SEQ ID No. 1156, SEQ ID No. 1157, SEQ ID No. 178, SEQ ID No. 179, SEQ ID No. 1158, SEQ ID No. 182, SEQ ID No. 183, SEQ ID No. 184, SEQ ID No. 185, SEQ ID No. 1159, SEQ ID No. 186, SEQ ID No. 1160, SEQ ID No. 187, SEQ ID No. 188, SEQ ID No. 189, SEQ ID

No. 190, SEQ ID No. 1161, SEQ ID No. 1162, SEQ ID No. 191, SEQ ID No. 192, SEQ ID No. 194, SEQ ID No. 195, SEQ ID No. 1163, SEQ ID No. 196, SEQ ID No. 201, SEQ ID No. 202, SEQ ID No. 209, SEQ ID No. 212, SEQ ID No. 221, SEQ ID No. 224, SEQ ID No. 1167, SEQ ID No. 226, SEQ ID No. 227, SEQ ID No. 228, SEQ ID No. 229, SEQ ID No. 230, SEQ ID No. 231, SEQ ID No. 232, SEQ ID No. 1169, SEQ ID No. 1170, SEQ ID No. 1171, SEQ ID No. 234, SEQ ID No. 235, SEQ ID No. 236, SEQ ID No. 1172, SEQ ID No. 243, SEQ ID No. 251, SEQ ID No. 252, SEQ ID No. 1176, SEQ ID No. 253, SEQ ID No. 255, SEQ ID No. 254, SEQ ID No. 256, SEQ ID No. 1177, SEQ ID No. 1178, SEQ ID No. 262, SEQ ID No. 263, SEQ ID No. 1264, SEQ ID No. 278, SEQ ID No. 279, SEQ ID No. 1180, SEQ ID No. 280, SEQ ID No. 290, SEQ ID No. 291, SEQ ID No. 292, SEQ ID No. 296, SEQ ID No. 1181, SEQ ID No. 297, SEQ ID No. 298, SEQ ID No. 300, SEQ ID No. 1265, SEQ ID No. 322, SEQ ID No. 324, SEQ ID No. 325, SEQ ID No. 370, SEQ ID No. 1186, SEQ ID No. 371, SEQ ID No. 372, SEQ ID No. 1187, SEQ ID No. 373, SEQ ID No. 378, SEQ ID No. 1266, SEQ ID No. 382, SEQ ID No. 383, SEQ ID No. 384, SEQ ID No. 385, SEQ ID No. 386, SEQ ID No. 1188, SEQ ID No. 1189, SEQ ID No. 391, SEQ ID No. 392, SEQ ID No. 398, SEQ ID No. 400, SEQ ID No. 403, SEQ ID No. 1191, SEQ ID No. 423, SEQ ID No. 435, SEQ ID No. 445, SEQ ID No. 450, SEQ ID No. 1193, SEQ ID No. 456, SEQ ID No. 460, SEQ ID No. 461, SEQ ID No. 465, SEQ ID No. 1196, SEQ ID No. 471, SEQ ID No. 473, SEQ ID No. 475, SEQ ID No. 481, SEQ ID No. 484, SEQ ID No. 487, SEQ ID No. 488, SEQ ID No. 489, SEQ ID No. 490, SEQ ID No. 491, SEQ ID No. 492, SEQ ID No. 493, SEQ ID No. 494, SEQ ID No. 495, SEQ ID No. 496, SEQ ID No. 497, SEQ ID No. 498, SEQ ID No. 499, SEQ ID No. 502, SEQ ID No. 1267, SEQ ID No. 1268, SEQ ID No. 508, SEQ ID No. 510, SEQ ID No. 509, SEQ ID No. 512, SEQ ID No. 515, SEQ ID No. 519, SEQ ID No. 1197, SEQ ID No. 521, SEQ ID No. 1198, SEQ ID No. 522, SEQ ID No. 524, SEQ ID No. 528, SEQ ID No. 534, SEQ ID No. 537, SEQ ID No. 1269, SEQ ID No. 1270, SEQ ID No. 548, SEQ ID No. 551, SEQ ID No. 557, SEQ ID No. 1201, SEQ ID No. 1203, SEQ ID No. 562, SEQ ID No. 566, SEQ ID No. 593, SEQ ID No. 595, SEQ ID No. 600, SEQ ID No. 1271, SEQ ID No. 604, SEQ ID No. 611, SEQ ID No. 612, SEQ ID No. 614, SEQ ID No. 616, SEQ ID No. 625, SEQ ID No. 627, SEQ ID No. 628, SEQ ID No. 629, SEQ ID No. 631, SEQ ID No. 641, SEQ ID No. 1272, SEQ ID No. 648, SEQ ID No. 1212, SEQ ID No. 663, SEQ ID No. 685, SEQ ID No. 707, SEQ ID No. 714, SEQ ID No. 715, SEQ ID No. 716, SEQ ID No. 717, SEQ ID No. 722, SEQ ID No. 746, SEQ ID No. 1273, SEQ ID No. 761, SEQ ID No. 764, SEQ ID No. 770, SEQ ID No. 1217, SEQ ID No. 783, SEQ ID No. 1274, SEQ ID No. 803, SEQ ID No. 815, SEQ ID No. 1220, SEQ ID No. 835, SEQ ID No. 1221, SEQ ID No. 844, SEQ ID No. 845, SEQ ID No. 846, SEQ ID No. 847, SEQ ID No. 848, SEQ ID No. 849, SEQ ID No. 850, SEQ ID No. 851, SEQ ID No. 1275, SEQ ID No. 852, SEQ ID No. 862, SEQ ID No. 1276, SEQ ID No. 1277, SEQ ID No. 873, SEQ ID No. 1223, SEQ ID No. 892, SEQ ID No. 919, SEQ ID No. 1225, SEQ ID No. 1278, SEQ ID No. 926, SEQ ID No. 1228, SEQ ID No. 1229, SEQ ID No. 1230, SEQ ID No. 1279, SEQ ID No. 1281, SEQ ID No. 1282, SEQ ID No. 1283, SEQ ID No. 948, SEQ ID No. 950, SEQ ID No. 949, SEQ ID No. 951, SEQ ID No. 980, SEQ ID No.

982, SEQ ID No. 1233, SEQ ID No. 999, SEQ ID No. 1000, SEQ ID No. 1001, SEQ ID No. 1002, SEQ ID No. 1008, SEQ ID No. 1285, SEQ ID No. 1235, SEQ ID No. 1016, SEQ ID No. 1019, SEQ ID No. 1027, SEQ ID No. 1036, SEQ ID No. 1241, SEQ ID No. 1048, SEQ ID No. 1049, SEQ ID No. 1050, SEQ ID No. 1053, SEQ ID No. 1054, SEQ ID No. 1064, SEQ ID No. 1076, SEQ ID No. 1091, SEQ ID No. 1288, SEQ ID No. 1093, SEQ ID No. 1289, SEQ ID No. 1101, SEQ ID No. 1103, SEQ ID No. 1245, SEQ ID No. 1246, SEQ ID No. 1247, SEQ ID No. 1290, SEQ ID No. 1291, SEQ ID No. 1115, SEQ ID No. 1116, SEQ ID No. 1118, SEQ ID No. 1120, SEQ ID No. 1249, SEQ ID No. 1121, SEQ ID No. 1250, SEQ ID No. 1126, SEQ ID No. 1251, SEQ ID No. 1127, SEQ ID No. 1128, SEQ ID No. 1130, SEQ ID No. 1129, SEQ ID No. 1131, SEQ ID No. 1136, SEQ ID No. 1253, SEQ ID No. 6844, SEQ ID No. 6846, SEQ ID No. 6847, SEQ ID No. 6848, and one of their representative fragments

Preferably, the invention relates to a polypeptide according to the invention, characterized in that it is a *Chlamydia pneumoniae* polypeptide or one of its representative fragments which is involved in the intermediate metabolism, in particular in the metabolism of sugars and/or of cofactors, and in that it is chosen from the polypeptides having the following sequences:

SEQ ID No. 2; SEQ ID No. 55; SEQ ID No. 56; SEQ ID No. 69; SEQ ID No. 75; SEQ ID No. 80; SEQ ID No. 100; SEQ ID No. 110; SEQ ID No. 114; SEQ ID No. 120; SEQ ID No. 121; SEQ ID No. 157; SEQ ID No. 160; SEQ ID No. 161; SEQ ID No. 172; SEQ ID No. 180; SEQ ID No. 181; SEQ ID No. 198; SEQ ID No. 200; SEQ ID No. 225; SEQ ID No. 248; SEQ ID No. 249; SEQ ID No. 276; SEQ ID No. 277; SEQ ID No. 318; SEQ ID No. 319; SEQ ID No. 320; SEQ ID No. 323; SEQ ID No. 331; SEQ ID No. 347; SEQ ID No. 375; SEQ ID No. 376; SEQ ID No. 381; SEQ ID No. 393; SEQ ID No. 394; SEQ ID No. 395; SEQ ID No. 396; SEQ ID No. 409; SEQ ID No. 446; SEQ ID No. 447; SEQ ID No. 448; SEQ ID No. 449; SEQ ID No. 513; SEQ ID No. 516; SEQ ID No. 571; SEQ ID No. 647; SEQ ID No. 662; SEQ ID No. 697; SEQ ID No. 718; SEQ ID No. 793; SEQ ID No. 794; SEQ ID No. 808; SEQ ID No. 809; SEQ ID No. 838; SEQ ID No. 839; SEQ ID No. 840; SEQ ID No. 853; SEQ ID No. 854; SEQ ID No. 918; SEQ ID No. 923; SEQ ID No. 929; SEQ ID No. 931; SEQ ID No. 938; SEQ ID No. 939; SEQ ID No. 958; SEQ ID No. 959; SEQ ID No. 960; SEQ ID No. 966; SEQ ID No. 995; SEQ ID No. 1021; SEQ ID No. 1040; SEQ ID No. 1041; SEQ ID No. 1042; SEQ ID No. 1085; SEQ ID No. 1100; SEQ ID No. 1102; SEQ ID No. 1117; SEQ ID No. 1118; SEQ ID No. 1119; SEQ ID No. 1120; SEQ ID No. 1135 and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, characterized in that it is a *Chlamydia pneumoniae* polypeptide or one of its representative fragments which is involved in the intermediate metabolism of nucleotides or nucleic acids, and in that it is chosen from the polypeptides having the following sequences:

SEQ ID No. 77; SEQ ID No. 78; SEQ ID No. 138; SEQ ID No. 189; SEQ ID No. 190; SEQ ID No. 233; SEQ ID No. 246; SEQ ID No. 338; SEQ ID No. 412; SEQ ID No. 421; SEQ ID No. 438;

SEQ ID No. 607; SEQ ID No. 648; SEQ ID No. 657; SEQ ID No. 740; SEQ ID No. 783; SEQ ID No. 967; SEQ ID No. 989; SEQ ID No. 990; SEQ ID No. 992; SEQ ID No. 1011; SEQ ID No. 1058; SEQ ID No. 1059; SEQ ID No. 1073; SEQ ID No. 1074 and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, characterized in that it is a *Chlamydia pneumoniae* polypeptide or one of its representative fragments which is involved in the metabolism of nucleic acids, and in that it is chosen from the polypeptides having the following sequences:

SEQ ID No. 14; SEQ ID No. 59; SEQ ID No. 70; SEQ ID No. 71; SEQ ID No. 97; SEQ ID No. 113; SEQ ID No. 137; SEQ ID No. 141; SEQ ID No. 169; SEQ ID No. 285; SEQ ID No. 287; SEQ ID No. 288; SEQ ID No. 313; SEQ ID No. 326; SEQ ID No. 358; SEQ ID No. 411; SEQ ID No. 443; SEQ ID No. 548; SEQ ID No. 569; SEQ ID No. 601; SEQ ID No. 651; SEQ ID No. 654; SEQ ID No. 658; SEQ ID No. 659; SEQ ID No. 664; SEQ ID No. 665; SEQ ID No. 694; SEQ ID No. 698; SEQ ID No. 704; SEQ ID No. 760; SEQ ID No. 762; SEQ ID No. 763; SEQ ID No. 786; SEQ ID No. 787; SEQ ID No. 788; SEQ ID No. 801; SEQ ID No. 802; SEQ ID No. 812; SEQ ID No. 819; SEQ ID No. 822; SEQ ID No. 870; SEQ ID No. 897; SEQ ID No. 898; SEQ ID No. 902; SEQ ID No. 908; SEQ ID No. 916; SEQ ID No. 954; SEQ ID No. 955; SEQ ID No. 961; SEQ ID No. 983; SEQ ID No. 996; SEQ ID No. 1007; SEQ ID No. 1012; SEQ ID No. 1013; SEQ ID No. 1014; SEQ ID No. 1015; SEQ ID No. 1038; SEQ ID No. 1137 and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, characterized in that it is a *Chlamydia pneumoniae* polypeptide or one of its representative fragments which is involved in the metabolism of amino acids or polypeptides, and in that it is chosen from the polypeptides having the following sequences:

SEQ ID No. 99; SEQ ID No. 111; SEQ ID No. 127; SEQ ID No. 134; SEQ ID No. 140; SEQ ID No. 174; SEQ ID No. 175; SEQ ID No. 176; SEQ ID No. 353; SEQ ID No. 377; SEQ ID No. 404; SEQ ID No. 523; SEQ ID No. 539; SEQ ID No. 559; SEQ ID No. 561; SEQ ID No. 586; SEQ ID No. 598; SEQ ID No. 609; SEQ ID No. 636; SEQ ID No. 687; SEQ ID No. 700; SEQ ID No. 701; SEQ ID No. 759; SEQ ID No. 790; SEQ ID No. 857; SEQ ID No. 861; SEQ ID No. 904; SEQ ID No. 936; SEQ ID No. 952; SEQ ID No. 962; SEQ ID No. 963; SEQ ID No. 964; SEQ ID No. 965; SEQ ID No. 991; SEQ ID No. 1003; SEQ ID No. 1004; SEQ ID No. 1005; SEQ ID No. 1018; SEQ ID No. 1067; SEQ ID No. 1110; SEQ ID No. 1111; SEQ ID No. 1112; SEQ ID No. 1114; SEQ ID No. 1121; SEQ ID No. 1122; SEQ ID No. 1123; SEQ ID No. 1124; SEQ ID No. 1125 and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, characterized in that it is a *Chlamydia pneumoniae* polypeptide or one of its representative fragments which is involved in the metabolism of polypeptides, and in that it is chosen from the polypeptides

having the following sequences:

SEQ ID No. 4; SEQ ID No. 44; SEQ ID No. 45; SEQ ID No. 48; SEQ ID No. 54; SEQ ID No. 112; SEQ ID No. 130; SEQ ID No. 155; SEQ ID No. 163; SEQ ID No. 212; SEQ ID No. 257; SEQ ID No. 307; SEQ ID No. 343; SEQ ID No. 405; SEQ ID No. 416; SEQ ID No. 458; SEQ ID No. 540; SEQ ID No. 541; SEQ ID No. 542; SEQ ID No. 543; SEQ ID No. 544; SEQ ID No. 560; SEQ ID No. 594; SEQ ID No. 652; SEQ ID No. 699; SEQ ID No. 723; SEQ ID No. 747; SEQ ID No. 817; SEQ ID No. 827; SEQ ID No. 871; SEQ ID No. 909; SEQ ID No. 910; SEQ ID No. 911; SEQ ID No. 912; SEQ ID No. 1023; SEQ ID No. 1051; SEQ ID No. 1052; SEQ ID No. 1081 and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, characterized in that it is a *Chlamydia pneumoniae* polypeptide or one of its representative fragments which is involved in the metabolism of fatty acids, and in that it is chosen from the polypeptides having the following sequences:

SEQ ID No. 76; SEQ ID No. 284; SEQ ID No. 308; SEQ ID No. 309; SEQ ID No. 310; SEQ ID No. 311; SEQ ID No. 312; SEQ ID No. 425; SEQ ID No. 433; SEQ ID No. 565; SEQ ID No. 688; SEQ ID No. 690; SEQ ID No. 691; SEQ ID No. 767; SEQ ID No. 797; SEQ ID No. 894; SEQ ID No. 895; SEQ ID No. 994; SEQ ID No. 1020; SEQ ID No. 1030; SEQ ID No. 1033; SEQ ID No. 1034; SEQ ID No. 1046; SEQ ID No. 1047; SEQ ID No. 1057 and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, characterized in that it is a *Chlamydia pneumoniae* polypeptide or one of its representative fragments which is involved in the synthesis of the wall, and in that it is chosen from the polypeptides having the following sequences:

SEQ ID No. 49; SEQ ID No. 50; SEQ ID No. 177; SEQ ID No. 178; SEQ ID No. 245; SEQ ID No. 610; SEQ ID No. 972; SEQ ID No. 974; SEQ ID No. 978; SEQ ID No. 1037 and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, characterized in that it is a *Chlamydia pneumoniae* polypeptide or one of its representative fragments which is involved in the transcription, translation and/or maturation process, and in that it is chosen from the polypeptides having the following sequences:

SEQ ID No. 90; SEQ ID No. 92; SEQ ID No. 131; SEQ ID No. 151; SEQ ID No. 199; SEQ ID No. 333; SEQ ID No. 334; SEQ ID No. 336; SEQ ID No. 379; SEQ ID No. 589; SEQ ID No. 590; SEQ ID No. 619; SEQ ID No. 630; SEQ ID No. 649; SEQ ID No. 739; SEQ ID No. 741; SEQ ID No. 806; SEQ ID No. 821; SEQ ID No. 843; SEQ ID No. 968; SEQ ID No. 971; SEQ ID No. 1061 and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, characterized in that it is a *Chlamydia pneumoniae* ribosomal polypeptide or one of its representative

fragments, and in that it is chosen from the polypeptides having the following sequences: SEQ ID No. 93; SEQ ID No. 94; SEQ ID No. 95; SEQ ID No. 136; SEQ ID No. 259; SEQ ID No. 332; SEQ ID No. 348; SEQ ID No. 583; SEQ ID No. 584; SEQ ID No. 588; SEQ ID No. 591; SEQ ID No. 592; SEQ ID No. 663; SEQ ID No. 666; SEQ ID No. 667; SEQ ID No. 669; SEQ ID No. 670; SEQ ID No. 671; SEQ ID No. 672; SEQ ID No. 673; SEQ ID No. 674; SEQ ID No. 675; SEQ ID No. 676; SEQ ID No. 677; SEQ ID No. 678; SEQ ID No. 679; SEQ ID No. 680; SEQ ID No. 681; SEQ ID No. 683; SEQ ID No. 684; SEQ ID No. 738; SEQ ID No. 781; SEQ ID No. 1008; SEQ ID No. 1024; SEQ ID No. 1025; SEQ ID No. 1066 and one of their representative fragments.

Preferably, the invention also relates to a polypeptide according to the invention, characterized in that it is a *Chlamydia pneumoniae* transport polypeptide or one of its representative fragments, and in that it is chosen from the polypeptides having the following sequences: SEQ ID No. 40; SEQ ID No. 41; SEQ ID No. 52; SEQ ID No. 105; SEQ ID No. 106; SEQ ID No. 107; SEQ ID No. 109; SEQ ID No. 133; SEQ ID No. 210; SEQ ID No. 211; SEQ ID No. 214; SEQ ID No. 215; SEQ ID No. 216; SEQ ID No. 217; SEQ ID No. 218; SEQ ID No. 219; SEQ ID No. 220; SEQ ID No. 223; SEQ ID No. 242; SEQ ID No. 260; SEQ ID No. 293; SEQ ID No. 299; SEQ ID No. 366; SEQ ID No. 369; SEQ ID No. 575; SEQ ID No. 602; SEQ ID No. 638; SEQ ID No. 639; SEQ ID No. 640; SEQ ID No. 643; SEQ ID No. 653; SEQ ID No. 702; SEQ ID No. 703; SEQ ID No. 724; SEQ ID No. 732; SEQ ID No. 855; SEQ ID No. 856; SEQ ID No. 901; SEQ ID No. 906; SEQ ID No. 933; SEQ ID No. 942; SEQ ID No. 1043; SEQ ID No. 1086; SEQ ID No. 1105 and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, characterized in that it is a *Chlamydia pneumoniae* polypeptide or one of its representative fragments which is involved in the virulence process, and in that it is chosen from the polypeptides having the following sequences: SEQ ID No. 546; SEQ ID No. 550; SEQ ID No. 778; SEQ ID No. 779; SEQ ID No. 886 and one of their representative fragments.

Preferably, the invention relates to a polypeptide according to the invention, characterized in that it is a *Chlamydia pneumoniae* polypeptide or one of its representative fragments which is involved in the secretory system and/or which is secreted, and in that it is chosen from the polypeptides having the following sequences: SEQ ID No. 751; SEQ ID No. 874; SEQ ID No. 875; SEQ ID No. 876; SEQ ID No. 883; SEQ ID No. 884; SEQ ID No. 885 and one of their representative fragments.

The secreted polypeptides, including the Type III and other, non-Type III secreted polypeptides, of the present invention, as well as the corresponding nucleotide sequences, may be detected by techniques known to persons skilled in the art, such as for example the techniques using cloning combined with vectors allowing the expression of the said polypeptides fused to export markers such as the *luc* gene for luciferase or the *PhoA* gene for alkaline phosphatase.

Preferably, the invention relates to a polypeptide according to the invention, characterized in that it is a polypeptide specific to *Chlamydia pneumoniae* or one of its representative fragments(with a Blast E value of $>10^{-5}$), and in that it is chosen from the polypeptides having the following sequences:

SEQ ID No. 7; SEQ ID No. 8; SEQ ID No. 17; SEQ ID No. 18; SEQ ID No. 19; SEQ ID No. 20; SEQ ID No. 22; SEQ ID No. 23; SEQ ID No. 24; SEQ ID No. 51; SEQ ID No. 60; SEQ ID No. 63; SEQ ID No. 65; SEQ ID No. 66; SEQ ID No. 67; SEQ ID No. 83; SEQ ID No. 84; SEQ ID No. 86; SEQ ID No. 87; SEQ ID No. 125; SEQ ID No. 143; SEQ ID No. 144; SEQ ID No. 179; SEQ ID No. 182; SEQ ID No. 184; SEQ ID No. 185; SEQ ID No. 187; SEQ ID No. 221; SEQ ID No. 252; SEQ ID No. 254;; SEQ ID No. 278; SEQ ID No. 279; SEQ ID No. 387; SEQ ID No. 388; SEQ ID No. 397; SEQ ID No. 1048; SEQ ID No. 1049; SEQ ID No. 1050; SEQ ID No. 1128; SEQ ID No. 1130; SEQ ID No. 1131 and one of their representative fragments.

In general, in the present invention, the functional group to which a polypeptide of the invention belongs, as well as its corresponding nucleotide sequence, may be determined either by comparative analogy with sequences already known, or by the use of standard techniques of biochemistry, of cytology combined with the techniques of genetic engineering such as immunoaffinity, localization by immunolabelling, differential extraction, measurement of enzymatic activity, study of the activity inducing or repressing expression or the study of expression in *E. coli*.

It is clearly understood, on the one hand, that, in the present invention, the nucleotide sequences (ORF) and the amino acid sequences (SEQ ID No. 2 to SEQ ID No. 1291 and SEQ ID No. 6844 to SEQ ID No. 6848) which are listed by functional group, are not exhaustive within the group considered. Moreover, it is also clearly understood that, in the present invention, a nucleotide sequence (ORF) or an amino acid sequence mentioned within a given functional group may also be part of another group taking into account, for example, the interrelationship between the groups listed. Accordingly, and as an example of this interrelationship, an exported and/or secreted polypeptide as well as its coding nucleotide sequence may also be involved in the *Chlamydia pneumoniae* virulence process by modifying the defense mechanism of the infected host cell, or a transmembrane polypeptide or its coding nucleotide sequence is also part of the polypeptides or coding nucleotide sequences of the cellular envelope.

The subject of the present invention is also the nucleotide and/or polypeptide sequences according to the invention, characterized in that the said sequences are recorded on a medium, called recording medium, whose type and nature facilitate the reading, the analysis and the exploitation of the said sequences. These media may of course also contain other information extracted from the present invention, such as in particular the analogies with already known sequences, such as those mentioned in Table 1 of the present description, and/or may contain, in addition, information relating to the nucleotide and/or polypeptide sequences of other microorganisms so as to facilitate the comparative analysis and the exploitation of the results obtained.

Among these recording media, computer-readable media, such as magnetic, optical, electrical and hybrid media such as, for example, floppy disks, CD-ROMs or recording cassettes, are preferred in particular.

The invention also relates to nucleotide sequences which can be used as primer or probe, characterized in that the said sequences are chosen from the nucleotide sequences according to the invention.

The invention relates, in addition, to the use of a nucleotide sequence according to the invention, as primer or probe, for the detection and/or amplification of nucleic acid sequences.

The nucleotide sequences according to the invention may thus be used to amplify nucleotide sequences, in particular by the PCR technique (polymerase chain reaction) (Erlich, 1989; Innis et al., 1990; Rolfs et al., 1991, and White et al., 1997).

These oligodeoxyribonucleotide or oligoribonucleotide primers correspond to representative nucleotide fragments, and are advantageously at least 8 nucleotides, preferably at least 12 nucleotides, 15 nucleotides and still more preferably at least 20 nucleotides long.

Other techniques for amplifying the target nucleic acid may be advantageously used as alternatives to PCR.

The nucleotide sequences of the invention, in particular the primers according to the invention, may also be used in other methods for amplifying a target nucleic acid, such as:

- the TAS (Transcription-based Amplification System) technique described by Kwoh et al. in 1989;
- the 3SR (Self-Sustained Sequence Replication) technique described by Guatelli et al. in 1990;
- the NASBA (Nucleic Acid Sequence Based Amplification) technique described by Kievitis et al. in 1991;
- the SDA (Strand Displacement Amplification) technique (Walker et al., 1992);
- the TMA (Transcription Mediated Amplification) technique.

The polynucleotides of the invention may also be used in techniques for amplifying or for modifying the nucleic acid serving as probe, such as:

- the LCR (Ligase Chain Reaction) technique described by Landegren et al. in 1988 and perfected by Barany et al. in 1991, which uses a thermostable ligase;
- the RCR (Repair Chain Reaction) technique described by Segev in 1992;
- the CPR (Cycling Probe Reaction) technique described by Duck et al. in 1990;
- the Q-beta-replicase amplification technique described by Miele et al. in 1983 and perfected in particular by Chu et al. in 1986, Lizardi et al. in 1988, and then by Burg et al. as well as by Stone et al. in 1996.

The invention also relates to the nucleotide sequences of fragments which can be obtained by amplification with the aid of at least one primer according to the invention. The present invention encompasses both hybridization probes and primers. In general, the complementary probes should be of a length sufficient to form a stable hybrid complex with the target sequences . Primers,

while complementary to the target sequences need not form stable hybridization complexes with the target sequences alone. Rather, primers form stable complexes with the target sequences in the presence of polymerase to permit extension of the primer.

In the case where the target polynucleotide to be detected is possibly an RNA, for example an mRNA, it will be possible to use, prior to the use of an amplification reaction with the aid of at least one primer according to the invention or to the use of a method of detection with the aid of at least one probe of the invention, a reverse transcriptase-type enzyme so as to obtain a cDNA from the RNA contained in the biological sample. The cDNA obtained will then serve as target for the primer(s) or the probe(s) used in the amplification or detection method according to the invention.

The detection probe will be chosen so that it hybridizes with the target sequence or the amplicon generated from the target sequence. Such a detection probe will advantageously have as sequence a sequence of at least 12 nucleotides, in particular of at least 20 nucleotides, and preferably at least 100 nucleotides.

The invention also comprises the nucleotide sequences which can be used as probe or primer according to the invention, characterized in that they are labelled with a radioactive compound or with a nonradioactive compound.

The nonlabelled nucleotide sequences may be used directly as probes or primers; however, the sequences are generally labelled with a radioactive element (^{32}P, ^{35}S, ^{3}H, ^{125}I) or with a nonradioactive molecule (biotin, acetylaminofluorene, digoxigenin, 5-bromo-deoxyuridine, fluorescein) so as to obtain probes which can be used in numerous applications.

Examples of nonradioactive labelling of nucleotide sequences are described, for example, in French patent No. 78,10975 or by Urdea et al. or by Sanchez-Pescador et al. in 1988.

In the latter case, one of the labelling methods described in patents FR-2 422 956 and FR-2 518 755 may also be used.

The invention also relates to the nucleotide sequences of fragments which can be obtained by hybridization with the aid of at least one probe according to the invention.

The hybridization technique may be performed in various ways (Matthews et al., 1988). The most common method consists in immobilizing the nucleic acid extracted from *Chlamydia pneumoniae* cells on a support (such as nitrocellulose, nylon, polystyrene) and in incubating, under well-defined conditions, the target nucleic acid immobilized with the probe. After the hybridization, the excess probe is removed and the hybrid molecules formed are detected by the appropriate method (measurement of the radioactivity, of the fluorescence or of the enzymatic activity linked to the probe).

The invention also comprises the nucleotide sequences according to the invention, characterized in that they are covalently or noncovalently immobilized on a support.

According to another advantageous embodiment of the nucleic sequences according to the invention, the latter may be used immobilized on a support and may thus serve to capture, through

specific hybridization, the target nucleic acid obtained from the biological sample to be tested. If necessary, the solid support is separated from the sample and the hybridization complex formed between the so-called capture probe and the target nucleic acid is then detected by means of a second probe, called detection probe, labelled with an easily detectable element.

The nucleotide sequences according to the invention may also be used in new analytical systems, DNA chips, which allow sequencing, the study of mutations and of the expression of genes, and which are currently of interest given their very small size and their high capacity in terms of number of analyses.

The principle of the operation of these chips is based on molecular probes, most often oligonucleotides, which are attached onto a miniaturized surface, generally of the order of a few square centimetres. During an analysis, a sample containing fragments of a target nucleic acid to be analysed, for example DNA or RNA labelled, for example, after amplification, is deposited onto the DNA chip in which the support has been coated beforehand with probes. Bringing the labelled target sequences into contact with the probes leads to the formation, through hybridization, of a duplex according to the rule of pairing defined by J.D. Watson and F. Crick. After a washing step, analysis of the surface of the chip allows the effective hybridizations to be located by means of the signals emitted by the labels tagging the target. A hybridization fingerprint results from this analysis which, by appropriate computer processing, will make it possible to determine information such as the presence of specific fragments in the sample, the determination of sequences and the presence of mutations.

The chip consists of a multitude of molecular probes, precisely organized or arrayed on a solid support whose surface is miniaturized. It is at the centre of a system where other elements (imaging system, microcomputer) allow the acquisition and interpretation of a hybridization fingerprint.

The hybridization supports are provided in the form of flat or porous surfaces (pierced with wells) composed of various materials. The choice of a support is determined by its physicochemical properties, or more precisely, by the relationship between the latter and the conditions under which the support will be placed during the synthesis or the attachment of the probes or during the use of the chip. It is therefore necessary, before considering the use of a particular support (R.S. Matson et al., 1994), to consider characteristics such as its stability to pH, its physical strength, its reactivity and its chemical stability as well as its capacity to nonspecifically bind nucleic acids. Materials such as glass, silicon and polymers are commonly used. Their surface is, in a first step, called "functionalization", made reactive towards the groups which it is desired to attach thereon. After the functionalization, so-called spacer molecules are grafted onto the activated surface. Used as intermediates between the surface and the probe, these molecules of variable size render unimportant the surface properties of the supports, which often prove to be problematic for the synthesis or the attachment of the probes and for the hybridization.

Among the hybridization supports, there may be mentioned glass which is used, for

example, in the method of in situ synthesis of oligonucleotides by photochemical addressing developed by the company Affymetrix (E.L. Sheldon, 1993), the glass surface being activated by silane. Genosensor Consortium (P. Mérel, 1994) also uses glass slides carrying wells 3 mm apart, this support being activated with epoxysilane.

Polymers or silicon may also be mentioned among these hybridization supports. For example, the Andrein Mirzabekov team has developed a chip consisting of polyacrylamide squares polymerized on a silanized glass surface (G. Yershov et al., 1996). Several teams use silicon, in particular the IFOS laboratory of Ecole Centrale of Lyon which uses a silicon semiconductor substrate which is p-doped by introducing it into its crystalline structure atoms whose valency is different from that of silicon. Various types of metals, in particular gold and platinum, may also be used as support (Genosensor Consortium (K. Beattie et al., 1993)).

The probes according to the invention may be synthesized directly in situ on the supports of the DNA chips. This in situ synthesis may be carried out by photochemical addressing (developed by the company Affymax (Amsterdam, Holland) and exploited industrially by its subsidiary Affymetrix (United States)) or based on the VLSIPS (very large scale immobilized polymer synthesis) technology (S.P.A. Fodor et al., 1991) which is based on a method of photochemically directed combinatory synthesis and the principle of which combines solid-phase chemistry, the use of photolabile protecting groups and photolithography.

The probes according to the invention may be attached to the DNA chips in various ways such as electrochemical addressing, automated addressing or the use of probe printers (T. Livache et al., 1994; G. Yershov et al., 1996; J. Derisi et al., 1996, and S. Borman, 1996).

The revealing of the hybridization between the probes of the invention, deposited or synthesized in situ on the supports of the DNA chips, and the sample to be analysed, may be determined, for example, by measurement of fluorescent signals, by radioactive counting or by electronic detection.

The use of fluorescent molecules such as fluorescein constitutes the most common method of labelling the samples. It allows direct or indirect revealing of the hybridization and allows the use of various fluorochromes.

Affymetrix currently provides an apparatus or a scanner designed to read its Gene Chip™ chips. It makes it possible to detect the hybridizations by scanning the surface of the chip in confocal microscopy (R.J. Lipshutz et al., 1995). Other methods of detecting fluorescent signals have been tested: coupling of an epifluorescence microscope and a CCD camera (G. Yershov et al., 1996), the use of an optical fibre collecting system (E.L. Sheldon, 1993). A conventional method consists in carrying out an end labelling, with phosphorus 32, of the target sequences, by means of an appropriate apparatus, the Phosphorimager (marketed by Molecular Dynamics). The electronic detection is based on the principle that the hybridization of two nucleic acid molecules is accompanied by physical phenomena which can be quantified under certain conditions (system developed by Ecole Centrale of

Lyon and called GEN-FET (GEN field effect transistor)). Genosensor Consortium and the company Beckman Instruments who are developing an electronic chip or Permittivity Chips™ may also be mentioned (K. Beattie et al., 1993).

The nucleotide sequences according to the invention may thus be used in DNA chips to carry out the analysis of mutations. This analysis is based on the production of chips capable of analysing each base of a nucleotide sequence according to the invention.

The nucleotide sequences according to the invention may also be used in DNA chips to carry out the analysis of the expression of the *Chlamydia pneumoniae* genes. This analysis of the expression of *Chlamydia pneumoniae* genes is based on the use of chips where probes of the invention, chosen for their specificity to characterize a given gene, are present (D.J. Lockhart et al., 1996; D.D. Shoemaker et al., 1996). For the methods of analysis of gene expression using the DNA chips, reference may, for example, be made to the methods described by D.J. Lockhart et al. (1996) and Sosnowsky et al. (1997) for the synthesis of probes in situ or for the addressing and the attachment of previously synthesized probes. The target sequences to be analysed are labelled and in general fragmented into sequences of about 50 to 100 nucleotides before being hybridized onto the chip. After washing as described, for example, by D.J. Lockhart et al. (1996) and application of different electric fields (Sosnowsky et al., 1997), the labelled compounds are detected and quantified, the hybridizations being carried out at least in duplicate. Comparative analyses of the signal intensities obtained with respect to the same probe for different samples and/or for different probes with the same sample, determine the differential expression of RNA or of DNA derived from the sample.

The nucleotide sequences according to the invention may, in addition, be used in DNA chips where other nucleotide probes specific for other microorganisms are also present, and may allow the carrying out of a serial test allowing rapid identification of the presence of a microorganism in a sample.

Accordingly, the subject of the invention is also the nucleotide sequences according to the invention, characterized in that they are immobilized on a support of a DNA chip.

The DNA chips, characterized in that they contain at least one nucleotide sequence according to the invention, immobilized on the support of the said chip, also form part of the invention.

The said chips will preferably contain several probes or nucleotide sequences of the invention of different length and/or corresponding to different genes so as to identify, with greater certainty, the specificity of the target sequences or the desired mutation in the sample to be analysed.

Accordingly, the analyses carried out by means of primers and/or probes according to the invention, immobilized on supports such as DNA chips, will make it possible, for example, to identify, in samples, mutations linked to variations such as intraspecies variations. These variations may be correlated or associated with pathologies specific to the variant identified and will make it possible to select the appropriate treatment.

The invention thus comprises a DNA chip according to the invention, characterized in that it contains, in addition, at least one nucleotide sequence of a microorganism different from *Chlamydia pneumoniae,* immobilized on the support of the said chip; preferably, the different microorganism will be chosen from an associated microorganism, a bacterium of the *Chlamydia* family, and a variant of the species *Chlamydia pneumoniae.*

Another subject of the present invention is a vector for the cloning and/or the expression of a sequence, characterized in that it contains a nucleotide sequence according to the invention. Among the said vectors according to the invention, the vectors containing a nucleotide sequence encoding a polypeptide of the cellular, preferably outer, envelope of *Chlamydia pneumoniae* or one of its representative fragments, are preferred. In a specific embodiment, the vectors contain a nucleotide sequence encoding a *Chlamydia pneumoniae* secreted polypeptide or one of its representative fragments or encoding a transport polypeptide, a surface exposed polypeptide, a lipoprotein or one of its representative fragments, a polypeptide involved in lipopolysaccharide (LPS) biosynthesis, a Type III and non-Type III secreted polypeptide, a polypeptide containing RGD attachment sites, a cell wall anchored surface polypeptide, a polypeptide not found in *Chlamydia trachomatis*, a ribosomal polypeptide or a polypeptide involved in secretion, transcription, translation, maturation of proteins, a polypeptide involved in the synthesis of the wall, a polypeptide involved in the virulence, a polypeptide involved in the intermediate metabolism, in particular in the metabolism of sugars and/or of cofactors, a polypeptide involved in the metabolism of nucleotides, of amino acids, of nucleic acids or of fatty acids of *Chlamydia pneumoniae* or one of their representative fragments, or a polypeptide specific to *Chlamydia pneumoniae.*

According to the invention, the vectors comprise the elements necessary to allow the expression and/or the secretion of the said nucleotide sequences in a given host cell, and form part of the invention. The vector should, in this case, comprise a promoter, signals for initiation and for termination of translation, as well as appropriate regions for regulation of transcription. It should be capable of being stably maintained in the host cell and may optionally possess particular signals specifying the secretion of the translated protein. These different elements are chosen according to the host cell used. To this effect, the nucleotide sequences according to the invention may be inserted into autonomously-replicating vectors within the chosen host, or integrative vectors in the chosen host.

Any of the standard methods known to those skilled in the art for the insertion of DNA fragments into a vector may be used to construct expression vectors containing a chimeric gene consisting of appropriate transcriptional/translational control signals and the protein coding sequences. These methods may include *in vitro* recombinant DNA and synthetic techniques and *in vivo* recombinants (genetic recombination).

Expression of a polypeptide, peptide or derivative, or analogs thereof encoded by a polynucleotide sequence in SEQ ID No. 1 or ORFs contained within SEQ ID No. 1 may be regulated by a second nucleic acid sequence so that the protein or peptide is expressed in a host transformed

with the recombinant DNA molecule. For example, expression of a protein or peptide may be controlled by any promoter/enhancer element known in the art. Promoters which may be used to control expression include, but are not limited to, the CMV promoter, the SV40 early promoter region (Bernoist and Chambon, 1981, Nature 290:304-310), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto, *et al.*, 1980, Cell 22:787-797), the herpes thymidine kinase promoter (Wagner *et al.*, 1981, Proc. Natl. Acad. Sci. U.S.A. 78:1441-1445), the regulatory sequences of the metallothionein gene (Brinster *et al.*, 1982, Nature 296:39-42); prokaryotic expression vectors such as the ∃-lactamase promoter (Villa-Kamaroff, *et al.*, 1978, Proc. Natl. Acad. Sci. U.S.A. 75:3727-3731), or the *tac* promoter (DeBoer, *et al.*, 1983, Proc. Natl. Acad. Sci. U.S.A. 80:21-25); see also "Useful proteins from recombinant bacteria" in Scientific American, 1980, 242:74-94; plant expression vectors comprising the nopaline synthetase promoter region (Herrera-Estrella *et al.*, 1983, Nature 303:209-213) or the cauliflower mosaic virus 35S RNA promoter (Gardner, *et al.*, 1981, Nucl. Acids Res. 9:2871), and the promoter of the photosynthetic enzyme ribulose biphosphate carboxylase (Herrera-Estrella *et al.*, 1984, Nature 310:115-120); promoter elements from yeast or other fungi such as the Gal 4 promoter, the ADC (alcohol dehydrogenase) promoter, PGK (phosphoglycerol kinase) promoter, alkaline phosphatase promoter, and the following animal transcriptional control regions, which exhibit tissue specificity and have been utilized in transgenic animals: elastase I gene control region which is active in pancreatic acinar cells (Swift *et al.*, 1984, Cell 38:639-646; Ornitz *et al.*, 1986, Cold Spring Harbor Symp. Quant. Biol. 50:399-409; MacDonald, 1987, Hepatology 7:425-515); insulin gene control region which is active in pancreatic beta cells (Hanahan, 1985, Nature 315:115-122), immunoglobulin gene control region which is active in lymphoid cells (Grosschedl *et al.*, 1984, Cell 38:647-658; Adames *et al.*, 1985, Nature 318:533-538; Alexander *et al.*, 1987, Mol. Cell. Biol. 7:1436-1444), mouse mammary tumor virus control region which is active in testicular, breast, lymphoid and mast cells (Leder *et al.*, 1986, Cell 45:485-495), albumin gene control region which is active in liver (Pinkert *et al.*, 1987, Genes and Devel. 1:268-276), alpha-fetoprotein gene control region which is active in liver (Krumlauf *et al.*, 1985, Mol. Cell. Biol. 5:1639-1648; Hammer *et al.*, 1987, Science 235:53-58; alpha 1-antitrypsin gene control region which is active in the liver (Kelsey *et al.*, 1987, Genes and Devel. 1:161-171), beta-globin gene control region which is active in myeloid cells (Mogram *et al.*, 1985, Nature 315:338-340; Kollias *et al.*, 1986, Cell 46:89-94; myelin basic protein gene control region which is active in oligodendrocyte cells in the brain (Readhead *et al.*, 1987, Cell 48:703-712); myosin light chain-2 gene control region which is active in skeletal muscle (Sani, 1985, Nature 314:283-286), and gonadotropic releasing hormone gene control region which is active in the hypothalamus (Mason *et al.*, 1986, Science 234:1372-1378).

The vectors according to the invention are, for example, vectors of plasmid or viral origin. In a specific embodiment, a vector is used that comprises a promoter operably linked to a protein or peptide-encoding a nucleic acid sequence in SEQ ID No. 1, or ORFs contained within SEQ ID No. 1, one or more origins of replication, and, optionally, one or more selectable markers (*e.g.*, an

antibiotic resistance gene). Expression vectors comprise regulatory sequences that control gene expression, including gene expression in a desired host cell. Preferred vectors for the expression of the polypeptides of the invention include the pET-type plasmid vectors (Promega) or pBAD plasmid vectors (Invitrogen). Furthermore, the vectors according to the invention are useful for transforming host cells so as to clone or express the nucleotide sequences of the invention.

Expression can also be achieved using targeted homologous recombination to activate *Chlamydia pneumoniae* genes present in the cloned genomic DNA. A heterologous regulatory element may be inserted into a stable cell line or cloned microorganism, such that it is operatively linked with an endogenous *Chlamydia pneumoniae* gene present in the cloned genome, using techniques, such as targeted homologous recombination, which are well known to those of skill in the art (*See, e.g.*, Chappel, U.S. Patent No. 4,215,051 and Skoultchi, WO 91/06667 each of which is incorporated herein in its entirety).

Expression vector/host cell systems containing inserts of polynucleotide sequences in SEQ ID No. 1 or ORFs within SEQ ID No. 1, which encode polypeptides, peptides or derivatives, or analogs thereof, can be identified by three general approaches: (a) nucleic acid hybridization, (b) presence or absence of "marker" gene functions, and (c) expression of inserted sequences. In the first approach, the presence of a polynucleotide sequence inserted in an expression vector can be detected by nucleic acid hybridization using probes comprising sequences that are homologous to an inserted polynucleotide sequence. In the second approach, the recombinant vector/host system can be identified and selected based upon the presence or absence of certain "marker" gene functions (*e.g.*, thymidine kinase activity, resistance to antibiotics, transformation phenotype, occlusion body formation in baculovirus, etc.) caused by the insertion of a polynucleotide sequence in the vector. For example, if the polynucleotide sequence in SEQ ID No. 1 or ORFs within SEQ ID No. 1 is inserted within the marker gene sequence of the vector, recombinants containing the insert can be identified by the absence of the marker gene function. In the third approach, recombinant expression vectors can be identified by assaying the product of the polynucleotide sequence expressed by the recombinant. Such assays can be based, for example, on the physical or functional properties of the expressed polypeptide in *in vitro* assay systems, *e.g.*, binding with antibody, promotion of cell proliferation.

Once a particular recombinant DNA molecule is identified and isolated, several methods known in the art may be used to propagate it. The clones identified may be introduced into an appropriate host cell by standard methods, such as for example lipofection, electroporation, and heat shock. Once a suitable host system and growth conditions are established, recombinant expression vectors can be propagated and prepared in quantity.

The invention also encompasses the host cells transformed by a vector according to the invention. These cells may be obtained by introducing into host cells a nucleotide sequence inserted into a vector as defined above, and then culturing the said cells under conditions allowing the replication and/or the expression of the transfected nucleotide sequence.

The host cell may be chosen from eukaryotic or prokaryotic systems, such as for example bacterial cells (Olins and Lee, 1993), but also yeast cells (Buckholz, 1993), as well as animal cells, in particular cultures of mammalian cells (Edwards and Aruffo, 1993), and in particular Chinese hamster ovary (CHO) cells, but also insect cells in which methods using baculoviruses for example may be used (Luckow, 1993).

Furthermore, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Expression from certain promoters can be elevated in the presence of certain inducers; thus, expression of the genetically engineered polypeptide may be controlled. Furthermore, different host cells have characteristic and specific mechanisms for the translational and post-translational processing and modification (*e.g.*, glycosylation, phosphorylation) of proteins. Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein expressed. For example, expression in a bacterial system can be used to produce an unglycosylated core protein product. Expression in yeast will produce a glycosylated product. Expression in mammalian cells can be used to ensure "native" glycosylation of a heterologous protein. Furthermore, different vector/host expression systems may effect processing reactions to different extents.

A preferred host cell for the expression of the proteins of the invention consists of prokaryotic cells, such as Gram$^-$ bacteria. A further preferred host cell according to the invention is a bacterium belonging to the *Chlamydia* family, more preferably belonging to the species *Chlamydia pneumoniae* or chosen from a microorganism associated with the species *Chlamydia pneumoniae.*

In other specific embodiments, the polypeptides, peptides or derivatives, or analogs thereof may be expressed as a fusion, or chimeric protein product (comprising the protein, fragment, analog, or derivative joined via a peptide bond to a heterologous protein sequence (of a different protein)). Such a chimeric product can be made by ligating the appropriate nucleic acid sequences encoding the desired amino acid sequences to each other by methods known in the art, in the proper coding frame, and expressing the chimeric product by methods commonly known in the art. Alternatively, such a chimeric product may be made by protein synthetic techniques, *e.g.*, by use of a peptide synthesizer.

Genomic sequences can be cloned and expressed as translational gene products (*i.e.*, peptides, polypeptides, and proteins) or transcriptional gene products (*i.e.*, antisense and ribozymes).

The invention further relates to the intracellular production of an antisense nucleic acid sequence of SEQ ID No. 1 by transcription from an exogenous sequence. For example, a vector can be introduced *in vivo* such that it is taken up by a cell, within which cell the vector or a portion thereof is transcribed, producing an antisense nucleic acid (RNA) of the invention. Such a vector would contain a sequence encoding an antisense nucleic acid. Such a vector can remain episomal or become chromosomally integrated, as long as it can be transcribed to produce the desired antisense RNA. Such vectors can be constructed by recombinant DNA technology methods standard in the art.

Vectors can be plasmid, viral, or others known in the art, used for replication and expression in mammalian cells. Expression of the sequence encoding the an antisense RNA can be by any promoter known in the art to act in mammalian, preferably human, cells. Such promoters can be inducible or constitutive. Such promoters include but are not limited to: the CMV promoter, the SV40 early promoter region (Bernoist and Chambon, 1981, Nature 290:304-310), the promoter contained in the 3N long terminal repeat of Rous sarcoma virus (Yamamoto *et al.*, 1980, Cell 22:787-797), the herpes thymidine kinase promoter (Wagner *et al.*, 1981, Proc. Natl. Acad. Sci. U.S.A. 78:1441-1445), the regulatory sequences of the metallothionein gene (Brinster *et al.*, 1982, Nature 296:39-42), etc.

In a specific embodiment, the antisense oligonucleotide comprises catalytic RNA, or a ribozyme (see, *e.g.*, PCT International Publication WO 90/11364, published October 4, 1990; Sarver *et al.*, 1990, Science 247:1222-1225). In another embodiment, the oligonucleotide is a 2N-0-methylribonucleotide (Inoue *et al.*, 1987, Nucl. Acids Res. 15:6131-6148), or a chimeric RNA-DNA analog (Inoue *et al.*, 1987, FEBS Lett. 215:327-330).

In another embodiment, the antisense nucleic acids of the invention comprise a sequence complementary to at least a portion of an RNA transcript of a polynucleotide sequence in SEQ ID No. 1. However, absolute complementarity, although preferred, is not required. A sequence "complementary to at least a portion of an RNA," as referred to herein, means a sequence having sufficient complementarity to be able to hybridize with the RNA, forming a stable duplex; in the case of double-stranded antisense nucleic acid sequence, a single strand of the duplex DNA may thus be tested, or triplex formation may be assayed. The ability to hybridize will depend on both the degree of complementarity and the length of the antisense nucleic acid. Generally, the longer the hybridizing nucleic acid, the more base mismatches with an RNA transcribed from SEQ ID No. 1 may contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine the melting point of the hybridized complex.

The invention also relates to the animals, except humans, comprising one of the above-described transformed cells according to the invention.

The production of transgenic animals according to the invention overexpressing one or more of the *Chlamydia pneumoniae* genes will be preferably carried out on rats, mice or rabbits according to methods well known to persons skilled in the art such as viral or nonviral transfections. The transgenic animals overexpressing one or more of the said genes may be obtained by transfection of multiple copies of the said genes under the control of a powerful promoter of a ubiquitous nature, or which is selective for one type of tissue. The transgenic animals may also be obtained by homologous recombination on embryonic stem cells, transfer of these stem cells to embryos, selection of the chimeras affected at the level of the reproductive lines, and growth of the said chimeras.

The transformed cells as well as the transgenic animals according to the invention can be used in methods of preparing the recombinant polypeptide.

It is now possible to produce recombinant polypeptides in a relatively large quantity by genetic engineering using the cells transformed with expression vectors according to the invention or using transgenic animals according to the invention.

The methods of preparing a polypeptide of the invention in recombinant form, characterized in that they use a vector and/or a cell transformed with a vector according to the invention and/or a transgenic animal comprising one of the said transformed cells according to the invention, are themselves included in the present invention.

Among the said methods of preparing a polypeptide of the invention in recombinant form, the methods of preparation using a vector, and/or a cell transformed with the said vector and/or a transgenic animal comprising one of the said transformed cells, containing a nucleotide sequence encoding a polypeptide of the cellular envelope of *Chlamydia pneumoniae* or one of its representative fragments, more preferably encoding a polypeptide of the outer cellular envelope of *Chlamydia pneumoniae* or one of its fragment, are preferred.

Among the said methods of preparing a polypeptide of the invention in recombinant form, the methods of preparation using a vector, and/or a cell transformed with the said vector and/or a transgenic animal comprising one of the said transformed cells, containing a nucleotide sequence encoding a *Chlamydia pneumoniae* secreted polypeptide or one of its representative fragments or encoding a transport polypeptide, a surface exposed polypeptide, a lipoprotein or one of its representative fragments, a polypeptide involved in lipopolysaccharide biosynthesis, a Type III or other secreted polypeptide, a polypeptide containing RGD attachment sites, a cell wall anchored surface polypeptide, a polypeptide not found in *Chlamydia trachomatis*, a ribosomal polypeptide or a polypeptide involved in secretion, transcription, translation, maturation of proteins, a polypeptide involved in the synthesis of the wall, a polypeptide involved in the virulence, a polypeptide involved in the intermediate metabolism, in particular in the metabolism of sugars and/or of cofactors, a polypeptide involved in the metabolism of nucleotides, of amino acids, of nucleic acids or of fatty acids of *Chlamydia pneumoniae* or one of their representative fragments, or a polypeptide specific to *Chlamydia pneumoniae*, are also preferred.

The recombinant polypeptides obtained as indicated above may be provided either in glycosylated or non-glycosylated form and may or may not have the natural tertiary structure.

A preferred variant consists in producing a recombinant polypeptide fused to a "carrier" protein (chimeric protein). The advantage of this system is that it allows a stabilization and a reduction in proteolysis of the recombinant product, an increase in solubility during renaturation in vitro and/or a simplification of purification when the fusion partner has affinity for a specific ligand.

More particularly, the invention relates to a method of preparing a polypeptide of the invention comprising the following steps:

a) culture of the transformed cells under conditions allowing the expression of a recombinant polypeptide having a nucleic acid sequence according to the invention;

b) where appropriate, recovery of the said recombinant polypeptide.

When the method of preparing a polypeptide of the invention uses a transgenic animal according to the invention, the recombinant polypeptide is then extracted from the said animal.

The subject of the invention is also a polypeptide capable of being obtained by a method of the invention as described above.

The invention also comprises a method of preparing a synthetic polypeptide, characterized in that it uses an amino acid sequence of polypeptides according to the invention.

The invention also relates to a synthetic polypeptide obtained by a method according to the invention.

Polypeptides according to the invention may also be prepared by conventional techniques in the field of peptide synthesis under conditions suitable to produce the polypeptides encoded by the polynucleotide of the invention. This synthesis may be carried out in and recovered from a homogeneous solution or on a solid phase.

For example, the synthesis technique in a homogeneous solution described by Houbenweyl in 1974 may be used.

This method of synthesis consists in successively condensing, in pairs, the successive amino acids in the required order, or in condensing amino acids and fragments previously formed and already containing several amino acids in the appropriate order, or alternatively several fragments thus previously prepared, it being understood that care will have been taken to protect beforehand all the reactive functional groups carried by these amino acids or fragments, with the exception of the amine functional groups of one and the carboxyl functional groups of the other or vice versa, which should normally take part in the formation of the peptide bonds, in particular after activation of the carboxyl functional group, according to methods well known in peptide synthesis.

According to another preferred technique of the invention, the one described by Merrifield is used.

To manufacture a peptide chain according to the Merrifield method, a highly porous polymer resin is used, onto which the first C-terminal amino acid of the chain is attached. This amino acid is attached onto a resin via its carboxyl group and its amine functional group is protected. The amino acids which will constitute the peptide chain are thus attached, one after another, onto the amine group, each time deprotected beforehand, of the portion of the peptide chain already formed, and which is attached to the resin. When the entire peptide chain desired is formed, the protecting groups are removed from the various amino acids constituting the peptide chain and the peptide is detached from the resin with the aid of an acid.

The invention relates, in addition, to hybrid (fusion) polypeptides having at least one polypeptide or one of its representative fragments according to the invention, and a sequence of a polypeptide capable of eliciting an immune response in humans or animals.

Advantageously, the antigenic determinant is such that it is capable of eliciting a humoral

and/or cellular response. An antigenic determinant may be identified by screening expression libraries of the *Chlamydia pneumoniae* genome with antibodies contained in the serum of patients infected with a bacterium belonging to the species *Chlamydia pneumoniae*. An antigenic determinant may comprise a polypeptide or one of its representative fragments according to the invention, in glycosylated form, used in order to obtain immunogenic compositions capable of inducing the synthesis of antibodies directed against multiple epitopes. The said polypeptides or their glycosylated fragments also form part of the invention.

These hybrid molecules may consist, in part, of a carrier molecule for polypeptides or for their representative fragments according to the invention, combined with a portion which may be immunogenic, in particular an epitope of the diphtheria toxin, the tetanus toxin, a hepatitis B virus surface antigen (patent FR 79 21811), the poliomyelitis virus VP1 antigen or any other viral or bacterial toxin or antigen.

The methods of synthesizing the hybrid molecules include the methods used in genetic engineering to construct hybrid nucleotide sequences encoding the desired polypeptide sequences. Reference may be advantageously made, for example, to the technique for producing genes encoding fusion proteins described by Minton in 1984.

The said hybrid nucleotide sequences encoding a hybrid polypeptide as well as the hybrid polypeptides according to the invention, characterized in that they are recombinant polypeptides obtained by the expression of the said hybrid nucleotide sequences, also form part of the invention.

The invention also comprises the vectors characterized in that they contain one of the said hybrid nucleotide sequences. The host cells transformed by the said vectors, the transgenic animals comprising one of the said transformed cells as well as the methods of preparing recombinant polypeptides using the said vectors, the said transformed cells and/or the said transgenic animals of course also form part of the invention.

The polypeptides according to the invention, the antibodies according to the invention described below and the nucleotide sequences according to the invention may advantageously be used in *in vitro* and/or *in vivo* methods for the detection and/or the identification of bacteria belonging to the species *Chlamydia pneumoniae*, in a biological sample (biological tissue or fluid) which is likely to contain them. These methods, depending on the specificity of the polypeptides, of the antibodies and of the nucleotide sequences according to the invention which will be used, may in particular detect and/or identify the bacterial variants belonging to the species *Chlamydia pneumoniae* as well as the associated microorganisms capable of being detected by the polypeptides, the antibodies and the nucleotide sequences according to the invention which will be chosen. It may, for example, be advantageous to choose a polypeptide, an antibody or a nucleotide sequence according to the invention, which is capable of detecting any bacterium of the Chlamydia family by choosing a polypeptide, an antibody and/or a nucleotide sequence according to the invention which is specific to the family or, on the contrary, it will be most particularly advantageous to target a variant of the

species *Chlamydia pneumoniae*, which is responsible, for example, for the induction or the worsening of pathologies specific to the targeted variant, by choosing a polypeptide, an antibody and/or a nucleotide sequence according to the invention which is specific to the said variant.

The polypeptides according to the invention may advantageously be used in a method for the detection and/or the identification of bacteria belonging to the species *Chlamydia pneumoniae* or to an associated microorganism, in a biological sample (biological tissue or fluid) which is likely to contain them, characterized in that it comprises the following steps:

a) bringing this biological sample into contact with a polypeptide or one of its representative fragments according to the invention (under conditions allowing an immunological reaction between the said polypeptide and the antibodies which may be present in the biological sample);

b) detecting the antigen-antibody complexes which may be formed.

Preferably, the biological sample consists of a fluid, for example a human or animal serum, blood or biopsies.

Any conventional procedure may be used to carry out such a detection of the antigen-antibody complexes which may be formed.

By way of example, a preferred method uses immunoenzymatic procedures based on the ELISA technique, immunofluorescence procedures or radioimmunological procedures (RIA), and the like.

Accordingly, the invention also relates to the polypeptides according to the invention, labelled with the aid of a suitable label such as a label of the enzymatic, fluorescent or radioactive type.

Such methods comprise, for example, the following steps:

- deposition of defined quantities of a polypeptide composition according to the invention into the wells of a microtitre plate,
- introduction, into the said wells, of increasing dilutions of serum, or of a different biological sample as defined above, which has to be analysed,
- incubation of the microplate,
- introduction, into the wells of the microtitre plate, of labelled antibodies directed against human or animal immunoglobulins, these antibodies having been labelled with the aid of an enzyme selected from those which are capable of hydrolyzing a substrate, thereby modifying the absorption of the radiation of the latter, at least at a defined wavelength, for example at 550 nm,
- detection, by comparison with a control, of the quantity of substrate hydrolyzed.

The invention also relates to a kit or set for the detection and/or the identification of bacteria belonging to the species *Chlamydia pneumoniae* or to an associated microorganism, characterized in that it comprises the following components:

- a polypeptide according to the invention,

- where appropriate, the reagents for constituting the medium appropriate for the immunological or specific reaction,
- the reagents allowing the detection of the antigen-antibody complexes produced by the immunological reaction between the polypeptide(s) of the invention and the antibodies which may be present in the biological sample, it being possible for these reagents also to carry a label, or to be capable of being recognized in turn by a labelled reagent, more particularly in the case where the polypeptide according to the invention is not labelled,
- where appropriate, a reference biological sample (negative control) free of antibodies recognized by a polypeptide according to the invention,
- where appropriate, a reference biological sample (positive control) containing a predetermined quantity of antibodies recognized by a polypeptide according to the invention.

According to the invention, the polypeptides, peptides, fusion proteins or other derivatives, or analogs thereof encoded by a polynucleotide sequence in SEQ ID No. 1, may be used as an immunogen to generate antibodies which immunospecifically bind such an immunogen. Such antibodies may include, but are not limited to, polyclonal and monoclonal antibodies, humanized or chimeric antibodies, single chain antibodies, Fab fragments, $F(ab')_2$ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above. In a specific embodiment, the antibody to a polypeptide, peptide or other derivative, or analog thereof encoded by a polynucleotide sequence in SEQ ID No. 1 is a bispecific antibody (see generally, *e.g.* Fanger and Drakeman, 1995, *Drug News and Perspectives* 8: 133-137). Such a bispecific antibody is genetically engineered to recognize both (1) an epitope and (2) one of a variety of "trigger" molecules, *e.g.* Fc receptors on myeloid cells, and CD3 and CD2 on T cells, that have been identified as being able to cause a cytotoxic T-cell to destroy a particular target. Such bispecific antibodies can be prepared either by chemical conjugation, hybridoma, or recombinant molecular biology techniques known to the skilled artisan.

Various procedures known in the art may be used for the production of polyclonal antibodies to a polypeptide, peptide or other derivative, or analog thereof encoded by a polynucleotide sequence in SEQ ID No. 1. For the production of antibody, various host animals can be immunized by injection with a polypeptide, or peptide or other derivative, or analog thereof, including but not limited to rabbits, mice, rats, etc. Various adjuvants, depending on the host species, may be used to increase the immunological response, including but not limited to Stimulon™ QS-21 (Aquila Biopharmaceuticals, Inc., Framingham, MA), MPL™ (3-O-deacylated monophosphoryl lipid A; RIBI ImmunoChem Research, Inc., Hamilton, MT), aluminum phosphate, IL-12 (Genetics Institute, Cambridge, MA), Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, BCG (bacille Calmette-Guerin), and corynebacterium parvum. Alternatively, polyclonal antibodies may be prepared by purifying, on an affinity column

onto which a polypeptide according to the invention has been previously attached, the antibodies contained in the serum of patients infected with a bacterium belonging to the species *Chlamydia pneumoniae*.

For preparation of monoclonal antibodies directed toward a polypeptide, peptide or other derivative, or analog, any technique which provides for the production of antibody molecules by continuous cell lines in culture may be used. For example, the hybridoma technique originally developed by Kohler and Milstein (1975, Nature 256:495-497), as well as the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole *et al.*, 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96). In an additional embodiment of the invention, monoclonal antibodies can be produced in germ-free animals utilizing technology described in PCT/US90/02545. In another embodiment of the invention, transgenic non-human animals can be used for the production of human antibodies utilizing technology described in WO 98/24893 and WO 96/33735. According to the invention, human antibodies may be used and can be obtained by using human hybridomas (Cote *et al.*, 1983, Proc. Natl. Acad. Sci. U.S.A. 80:2026-2030) or by transforming human B cells with EBV virus *in vitro* (Cole *et al.*, 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, pp. 77-96). In fact, according to the invention, techniques developed for the production of "chimeric antibodies" (Morrison *et al.*, 1984, PROC. NATL. ACAD. SCI. U.S.A. 81:6851-6855; Neuberger *et al.*, 1984, Nature 312:604-608; Takeda *et al.*, 1985, Nature 314:452-454) by splicing the genes from a mouse antibody molecule specific for a polypeptide, peptide or other derivative, or analog together with genes from a human antibody molecule of appropriate biological activity can be used; such antibodies are within the scope of this invention.

According to the invention, techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce polypeptide or peptide-specific single chain antibodies. An additional embodiment of the invention utilizes the techniques described for the construction of Fab expression libraries (Huse *et al.*, 1989, Science 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity for polypeptides, derivatives, or analogs.

Antibody fragments which contain the idiotype of the molecule can be generated by known techniques. For example, such fragments include but are not limited to: the $F(ab')_2$ fragment which can be produced by pepsin digestion of the antibody molecule; the Fab' fragments which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragment, the Fab fragments which can be generated by treating the antibody molecule with papain and a reducing agent, and Fv fragments.

In addition, techniques have been developed for the production of chimerized (*See*, e.g., Boss, M. et al., U.S. Patent No. 4,816,397; and Cabilly, S. et al., U.S. Patent No. 5,585,089 each of which is incorporated herein by reference in its entirety) humanized antibodies (*See*, e.g., Queen, U.S. Patent No. 5,585,089, which is incorporated herein by reference in its entirety.) An immunoglobulin

light or heavy chain variable region consists of a "framework" region interrupted by three hypervariable regions, referred to as complementarily determining regions (CDRs). The extent of the framework region and CDRs have been precisely defined (See, "Sequences of Proteins of Immunological Interest", Kabat, E. et al., U.S. Department of Health and Human Services (1983). Briefly, humanized antibodies are antibody molecules from non-human species having one or more CDRs from the non-human species and a framework from a human immunoglobulin molecule.

The antibodies of the invention may also be labelled in the same manner as described above for the nucleic probes of the invention such as an enzymatic, fluorescent or radioactive type labelling.

The invention relates, in addition, to a method for the detection and/or the identification of bacteria belonging to the species *Chlamydia pneumoniae* or to an associated microorganism in a biological sample, characterized in that it comprises the following steps:

a) bringing the biological sample (biological tissue or fluid) into contact with a mono- or polyclonal antibody according to the invention (under conditions allowing an immunological reaction between the said antibodies and the polypeptides of the bacterium belonging to the species *Chlamydia pneumoniae* or to an associated microorganism which may be present in the biological sample, that is, under conditions suitable for the formation of immune complexes);
b) detecting the antigen-antibody complex which may be formed.

Also falling within the scope of the invention is a kit or set for the detection and/or the identification of bacteria belonging to the species *Chlamydia pneumoniae* or to an associated microorganism, characterized in that it comprises the following components:

- a polyclonal or monoclonal antibody according to the invention, labeled where appropriate;
- where appropriate, a reagent for constituting the medium appropriate for carrying out the immunological reaction;
- a reagent allowing the detection of the antigen-antibody complexes produced by the immunological reaction, it being possible for this reagent also to carry a label, or to be capable of being recognized in turn by a labelled reagent, more particularly in the case where the said monoclonal or polyclonal antibody is not labelled;
- where appropriate, reagents for carrying out the lysis of the cells in the sample tested.

The principle of the DNA chip which was explained above may also be used to produce protein "chips" on which the support has been coated with a polypeptide or an antibody according to the invention, or arrays thereof, in place of the DNA. These protein "chips" make it possible, for example, to analyze the biomolecular interactions (BIA) induced by the affinity capture of target analytes onto a support coated, for example, with proteins, by surface plasma resonance (SPR). Reference may be made, for example, to the techniques for coupling proteins onto a solid support which are described in EP 524 800 or to the methods describing the use of biosensor-type protein

chips such as the BIAcore-type technique (Pharmacia) (Arlinghaus et al., 1997, Krone et al., 1997, Chatelier et al., 1995). These polypeptides or antibodies according to the invention, capable of specifically binding antibodies or polypeptides derived from the sample to be analysed, may thus be used in protein chips for the detection and/or the identification of proteins in samples. The said protein chips may in particular be used for infectious diagnosis and may preferably contain, per chip, several polypeptides and/or antibodies of the invention of different specificity, and/or polypeptides and/or antibodies capable of recognizing microorganisms different from *Chlamydia pneumoniae*.

Accordingly, the subject of the present invention is also the polypeptides and the antibodies according to the invention, characterized in that they are immobilized on a support, in particular of a protein chip.

The protein chips, characterized in that they contain at least one polypeptide or one antibody according to the invention immobilized on the support of the said chip, also form part of the invention.

The invention comprises, in addition, a protein chip according to the invention, characterized in that it contains, in addition, at least one polypeptide of a microorganism different from *Chlamydia pneumoniae* or at least one antibody directed against a compound of a microorganism different from *Chlamydia pneumoniae*, immobilized on the support of the said chip.

The invention also relates to a kit or set for the detection and/or the identification of bacteria belonging to the species *Chlamydia pneumoniae* or to an associated microorganism, or for the detection and/or the identification of a microorganism characterized in that it comprises a protein chip according to the invention.

The subject of the present invention is also a method for the detection and/or the identification of bacteria belonging to the species *Chlamydia pneumoniae* or to an associated microorganism in a biological sample, characterized in that it uses a nucleotide sequence according to the invention.

More particularly, the invention relates to a method for the detection and/or the identification of bacteria belonging to the species *Chlamydia pneumoniae* or to an associated microorganism in a biological sample, characterized in that it comprises the following steps:

a) where appropriate, isolation of the DNA from the biological sample to be analysed, or optionally production of a cDNA from the RNA in the biological sample;
b) specific amplification of the DNA of bacteria belonging to the species *Chlamydia pneumoniae* or to an associated microorganism with the aid of at least one primer according to the invention;
c) detection of the amplification products.

These may be detected, for example, by the molecular hybridization technique using a nucleic probe according to the invention. This probe will be advantageously labelled with a nonradioactive (cold probe) or radioactive element.

For the purposes of the present invention, "DNA in the biological sample" or "DNA contained in the biological sample" will be understood to mean either the DNA present in the biological sample considered, or optionally the cDNA obtained after the action of a reverse transcriptase-type enzyme on the RNA present in the said biological sample.

Another aim of the present invention consists in a method according to the invention, characterized in that it comprises the following steps:

a) bringing a nucleotide probe according to the invention into contact with a biological sample, the DNA contained in the biological sample having, where appropriate, been previously made accessible to hybridization, under conditions allowing the hybridization of the probe to complementary base pairs of the DNA of a bacterium belonging to the species *Chlamydia pneumoniae* or to an associated microorganism;

b) detecting the hybridization complex formed between the nucleotide probe and the DNA in the biological sample.

The present invention also relates to a method according to the invention, characterized in that it comprises the following steps:

a) bringing a nucleotide probe immobilized on a support according to the invention into contact with a biological sample, the DNA in the sample having, where appropriate, been previously made accessible to hybridization, under conditions allowing the hybridization of the probe to the DNA of a bacterium belonging to the species *Chlamydia pneumoniae* or to an associated microorganism;

b) bringing the hybrid formed between the nucleotide probe immobilized on a support and the DNA contained in the biological sample, where appropriate after removal of the DNA in the biological sample which has not hybridized with the probe, into contact with a labelled nucleotide probe according to the invention;

c) detecting the new hybrid formed in step b).

According to an advantageous embodiment of the method for the detection and/or the identification defined above, it is characterized in that, prior to step a), the DNA in the biological sample is primer-extended and/or amplified beforehand with the aid of at least one primer according to the invention.

The invention relates, in addition, to a kit or set for the detection and/or the identification of bacteria belonging to the species *Chlamydia pneumoniae* or to an associated microorganism, characterized in that it comprises the following components:

a) a nucleotide probe according to the invention;

b) where appropriate, the reagents necessary for carrying out a hybridization reaction;

c) where appropriate, at least one primer according to the invention as well as the reagents (<u>e.g.</u>, polymerase and/or deoxynucleotide triphosphates) necessary for a DNA amplification reaction.

The invention also relates to a kit or set for the detection and/or the identification of bacteria belonging to the species *Chlamydia pneumoniae* or to an associated microorganism, characterized in that it comprises the following components:

a) a nucleotide probe, called capture probe, according to the invention;
b) an oligonucleotide probe, called detection probe, according to the invention;
c) where appropriate, at least one primer according to the invention as well as the reagents (e.g., polymerase and/or deoxynucleotide triphosphates) necessary for a DNA amplification reaction.

The invention also relates to a kit or set for the detection and/or the identification of bacteria belonging to the species *Chlamydia pneumoniae* or to an associated microorganism, characterized in that it comprises the following components:

a) at least one primer according to the invention;
b) where appropriate, the reagents necessary for carrying out a DNA amplification reaction;
c) where appropriate, a component which makes it possible to check the sequence of the amplified fragment, more particularly an oligonucleotide probe according to the invention.

The invention relates, in addition, to a kit or set for the detection and/or the identification of bacteria belonging to the species *Chlamydia pneumoniae* or to an associated microorganism, or for the detection and/or the identification of a microorganism characterized in that it comprises a DNA chip according to the invention.

The invention also relates to a method or to a kit or set according to the invention for the detection and/or the identification of bacteria belonging to the species *Chlamydia pneumoniae*, characterized in that the said primer and/or the said probe according to the invention are chosen from the nucleotide sequences specific to the species *Chlamydia pneumoniae*, in that the said polypeptides according to the invention are chosen from the polypeptides specific to the species *Chlamydia pneumoniae* and in that the said antibodies according to the invention are chosen from the antibodies directed against the polypeptides according to the invention chosen from the polypeptides specific to the species *Chlamydia pneumoniae*.

Preferably, the said method or the said kit or set above according to the invention, for the detection and/or the identification of bacteria belonging to the species *Chlamydia pneumoniae* is characterized in that the said primer and/or the said probe or the said polypeptides are chosen from the nucleotide sequences or polypeptides according to the invention which have been identified as being specific to the species *Chlamydia pneumoniae* and in that the said antibodies according to the invention are chosen from the antibodies directed against the polypeptides according to the invention chosen from the polypeptides identified as being specific to the species *Chlamydia pneumoniae*.

The invention relates, in addition, to a method or a kit or set according to the invention for the diagnosis of predispositions to, or of a condition caused by, cardiovascular diseases, preferably linked to the presence of atheroma, which are induced or worsened by a *Chlamydia pneumoniae*

infection.

The invention also relates to a method or a kit or set according to the invention for the diagnosis of predispositions to, or of conditions caused by, respiratory diseases induced or worsened by a *Chlamydia pneumoniae* infection; preferably, the said respiratory disease is asthma.

According to another aspect, the subject of the invention is the use of polypeptides according to the invention, of cells transformed with a vector according to the invention and/or of transformed animals according to the invention, for the biosynthesis or the biodegradation of organic or inorganic compounds.

As has been mentioned above, the nucleotide sequences of the invention were identified by homology with sequences known to encode, for example, polypeptides or fragments of enzymatic polypeptides involved in the biosynthesis or the biodegradation of organic or inorganic molecules.

It is thus possible to use the said polypeptides of the invention in a similar manner for the biosynthesis or the biodegradation of organic or inorganic compounds of industrial or therapeutic interest (called compounds of interest).

Among these polypeptides, there may be mentioned in particular the enzymes involved in metabolism, such as the proteolytic enzymes, amino transferases, glucose metabolism, or the enzymes which may be used in the biosynthesis of sugars, amino acids, fatty acids, polypeptides, nucleotides, nucleic acids or any other organic or inorganic compound or in the biodegradation of organic or inorganic compounds.

Among these polypeptides, there may be mentioned, in addition, the mutated or modified enzymes corresponding to mutated or modified polypeptides according to the invention which may also be used for the biosynthesis or the biodegradation of organic or inorganic compounds at the industrial level, such as, for example, the production of compounds of interest, the reprocessing of manufacturing residues applied to the food industries, to the papermaking industry or to the chemical and pharmaceutical industries.

The methods of biosynthesis or biodegradation of organic or inorganic compounds, characterized in that they use a polypeptide or one of its representative fragments according to the invention, transformed cells according to the invention and/or a transformed animal according to the invention, also form part of the invention.

The invention relates, in addition, to the use of a nucleotide sequence according to the invention, of a polypeptide according to the invention, of an antibody according to the invention, of a cell according to the invention, and/or of a transformed animal according to the invention, for the selection of an organic or inorganic compound capable of modulating, regulating, inducing or inhibiting the expression of genes, and/or of modifying the cellular replication of eukaryotic or prokaryotic cells or capable of inducing, inhibiting or worsening the pathologies linked to an infection by *Chlamydia pneumoniae* or one of its associated microorganisms.

The invention also comprises screening assays that comprise methods of selecting

compounds capable of binding to a polypeptide, fusion polypeptide or one of its representative fragments according to the invention, capable of binding to a nucleotide sequence according to the invention, or capable of recognizing an antibody according to the invention, and/or capable of modulating, regulating, inducing or inhibiting the expression of genes, and/or of modifying the growth or the cellular replication of eukaryotic or prokaryotic cells, or capable of inducing, inhibiting or worsening, in an animal or human organism, the pathologies linked to an infection by *Chlamydia pneumoniae* or one of its associated microorganisms, characterized in that it comprises the following steps:

a) bringing the said compound into contact with the said polypeptide, the said nucleotide sequence, with a transformed cell according to the invention and/or administering the said compound to a transformed animal according to the invention;

b) determining the capacity of the said compound to bind with the said polypeptide or the said nucleotide sequence, or to modulate, regulate, induce or inhibit the expression of genes, or to modulate growth or cellular replication, or to induce, inhibit or worsen in the said transformed animal, the pathologies linked to an infection by *Chlamydia pneumoniae* or one of its associated microorganisms.

The transformed cells and/or animals according to the invention may advantageously serve as a model and may be used in methods for studying, identifying and/or selecting compounds capable of being responsible for pathologies induced or worsened by *Chlamydia pneumoniae,* or capable of preventing and/or of treating these pathologies such as, for example, cardiovascular or respiratory diseases. In particular, the transformed host cells, in particular bacteria of the *Chlamydia* family whose transformation with a vector according to the invention may, for example, increase or inhibit its infectivity, or modulate the pathologies usually induced or worsened by the infection, may be used to infect animals in which the onset of pathologies will be monitored. These nontransformed animals, infected for example with transformed *Chlamydia* bacteria, may serve as a study model. In the same manner, the transformed animals according to the invention may, for example, exhibit predispositions to cardiovascular and/or respiratory diseases and thus be used in methods for selecting compounds capable of preventing and/or of treating the said diseases. The said methods using the said transformed cells and/or transformed animals form part of the invention.

The compounds capable of being selected may be organic compounds such as polypeptides or carbohydrates or any other organic or inorganic compounds already known, or new organic compounds produced using molecular modeling techniques and obtained by chemical or biochemical synthesis, these techniques being known to persons skilled in the art.

The said selected compounds may be used to modulate the growth and/or the cellular replication of *Chlamydia pneumoniae* or any other associated microorganism and thus to control infection by these microorganisms. The said compounds according to the invention may also be used to modulate the growth and/or the cellular replication of all eukaryotic or prokaryotic cells, in

particular tumour cells and infectious microorganisms, for which the said compounds will prove active, the methods which make it possible to determine the said modulations being well known to persons skilled in the art.

Compound capable of modulating the growth of a microorganism is understood to designate any compound which makes it possible to act, to modify, to limit and/or to reduce the development, the growth, the rate of proliferation and/or the viability of the said microorganism.

This modulation may be achieved, for example, by an agent capable of binding to a protein and thus of inhibiting or of potentiating its biological activity, or capable of binding to a membrane protein of the outer surface of a microorganism and of blocking the penetration of the said microorganism into the host cell or of promoting the action of the immune system of the infected organism directed against the said microorganism. This modulation may also be achieved by an agent capable of binding to a nucleotide sequence of a DNA or RNA of a microorganism and of blocking, for example, the expression of a polypeptide whose biological or structural activity is necessary for the growth or for the reproduction of the said microorganism.

Associated microorganism is understood to designate in the present invention any microorganism whose gene expression may be modulated, regulated, induced or inhibited, or whose growth or cellular replication may also be modulated by a compound of the invention. Associated microorganism is also understood to designate in the present invention any microorganism containing nucleotide sequences or polypeptides according to the invention. These microorganisms may, in some cases, contain polypeptides or nucleotide sequences identical or homologous to those of the invention may also be detected and/or identified by the detection and/or identification methods or kit according to the invention and may also serve as a target for the compounds of the invention.

The invention relates to the compounds capable of being selected by a method of selection according to the invention.

The invention also relates to a pharmaceutical composition comprising a compound chosen from the following compounds:

a nucleotide sequence according to the invention;

a polypeptide according to the invention;

a vector according to the invention;

an antibody according to the invention; and

a compound capable of being selected by a method of selection according to the invention, optionally in combination with a pharmaceutically acceptable vehicle.

An effective quantity is understood to designate a sufficient quantity of the said compound or antibody, or of a polypeptide of the invention, which makes it possible to modulate the growth of *Chlamydia pneumoniae* or of an associated microorganism.

The invention also relates to a pharmaceutical composition comprising one or more polypeptides according to the invention and/or one or more fusion polypeptides according to the

invention. Such compositions further comprise a pharmaceutically acceptable carrier or vehicle. Pharmaceutical compositions include compositions that comprise a polypeptide or fusion polypeptide that immunoreacts with seropositive serum of an individual infected with *Chlamydia pneumoniae*. In one embodiment, a pharmaceutical composition according to the invention can be utilized for the prevention or the treatment of an infection by a bacterium belonging to the species *Chlamydia pneumoniae* or by an associated microorganism.

The invention relates, in addition, to an immunogenic composition or a vaccine composition, characterized in that it comprises one or more polypeptides according to the invention and/or one or more hybrid (fusion) polypeptides according to the invention. Such compositions further comprise a pharmaceutically acceptable carrier or vehicle. Immunogenic compositions or fusion polypeptide include compositions that comprise a polypeptide that immunoreacts with seropositive serum of an individual infected with *Chlamydia pneumoniae*.

Immunogenic or vaccine compositions can also comprise DNA immunogenic or vaccine compositions comprising polynucleotide sequences of the invention operatively associated with a regulatory sequence that controls gene expression. Such compositions can include compositions that direct expression of a neutralizing epitope of *Chlamydia pneumoniae*.

The invention also comprises the use of a transformed cell according to the invention, for the preparation of a vaccine composition.

The invention also relates to a vaccine composition, characterized in that it contains a nucleotide sequence according to the invention, a vector according to the invention and/or a transformed cell according to the invention.

The invention also relates to the vaccine compositions according to the invention, for the prevention or the treatment of an infection by a bacterium belonging to the species *Chlamydia pneumoniae* or by an associated microorganism.

The invention also relates to the use of DNA encoding polypeptides of *Chlamydia pneumoniae*, in particular antigenic determinants, to be formulated as vaccine compositions. In accordance with this aspect of the invention, the DNA of interest is engineered into an expression vector under the control of regulatory elements, which will promote expression of the DNA, *i.e.*, promoter or enhancer elements. In one preferred embodiment, the promoter element may be cell-specific and permit substantial transcription of the DNA only in predetermined cells. The DNA may be introduced directly into the host either as naked DNA (U.S. Patent No. 5,679,647 incorporated herein by reference in their entirety) or formulated in compositions with other agents which may facilitate uptake of the DNA including viral vectors, *i.e.*, adenovirus vectors, or agents which facilitate immunization, such as bupivicaine and other local anesthetics (U.S. Patent 5,593,972 incorporated herein by reference in their entirety), saponins (U.S. Patent 5,739,118 incorporated herein by reference in their entirety) and cationic polyamines (published international application WO 96/10038 incorporated herein by reference in their entirety).

The DNA sequence encoding the antigenic polypeptide and regulatory element may be inserted into a stable cell line or cloned microorganism, using techniques, such as targeted homologous recombination, which are well known to those of skill in the art, and described *e.g.*, in Chappel, U.S. Patent No. 4,215,051; Skoultchi, WO 91/06667 each of which is incorporated herein by reference in its entirety.

Such cell lines and microorganisms may be formulated for vaccine purposes. In yet another embodiment, the DNA sequence encoding the antigenic polypeptide and regulatory element may be delivered to a mammalian host and introduced into the host genome via homologous recombination (*See*, Chappel, U.S. Patent No. 4,215,051; Skoultchi, WO 91/06667 each of which is incorporated herein by reference in its entirety.

Preferably, the immunogenic and/or vaccine compositions according to the invention intended for the prevention and/or the treatment of an infection by *Chlamydia pneumoniae* or by an associated microorganism will be chosen from the immunogenic and/or vaccine compositions comprising a polypeptide or one of its representative fragments corresponding to a protein, or one of its representative fragments, of the cellular envelope of *Chlamydia pneumoniae*. The vaccine compositions comprising nucleotide sequences will also preferably comprise nucleotide sequences encoding a polypeptide or one of its representative fragments corresponding to a protein, or one of its representative fragments, of the cellular envelope of *Chlamydia pneumoniae.*

Among these preferred immunogenic and/or vaccine compositions, the most preferred are those comprising a polypeptide or one of its representative fragments, or a nucleotide sequence or one of its representative fragments whose sequences are chosen from the nucleotide or amino acid sequences identified in this functional group and listed above.

The polypeptides of the invention or their representative fragments entering into the immunogenic compositions according to the invention may be selected by techniques known to persons skilled in the art, such as for example on the capacity of the said polypeptides to stimulate T cells, which results, for example, in their proliferation or the secretion of interleukins, and which leads to the production of antibodies directed against the said polypeptides.

In mice, in which a weight dose of the vaccine composition comparable to the dose used in humans is administered, the antibody reaction is tested by collecting serum followed by a study of the formation of a complex between the antibodies present in the serum and the antigen of the vaccine composition, according to the customary techniques.

According to the invention, the said vaccine compositions will be preferably in combination with a pharmaceutically acceptable vehicle and, where appropriate, with one or more appropriate immunity adjuvants.

Various types of vaccines are currently available for protecting humans against infectious diseases: attenuated live microorganisms (*M. bovis* - BCG for tuberculosis), inactivated microorganisms (influenza virus), acellular extracts (*Bordetella pertussis* for whooping cough),

recombinant proteins (hepatitis B virus surface antigen), polysaccharides (pneumococci). Experiments are underway on vaccines prepared from synthetic peptides or from genetically modified microorganisms expressing heterologous antigens. Even more recently, recombinant plasmid DNAs carrying genes encoding protective antigens were proposed as an alternative vaccine strategy. This type of vaccination is carried out with a particular plasmid derived from an *E. coli* plasmid which does not replicate *in vivo* and which encodes only the vaccinal protein. Animals were immunized by simply injecting the naked plasmid DNA into the muscle. This technique leads to the expression of the vaccine protein *in situ* and to a cell-type (CTL) and a humoral type (antibody) immune response. This double induction of the immune response is one of the main advantages of the technique of vaccination with naked DNA.

The vaccine compositions of the present invention can be evaluated in *in vitro* and *in vivo* animal models prior to host, e.g., human, administration. For example, *in vitro* neutralization assays such as those described by Peterson et al. (1988) can be utilized. The assay described by Peterson et al. (1988) is suitable for testing vaccine compositions directed toward either *Chlamydia pneumoniae* or *Chlamydia trachomatis*.

Briefly, hyper-immune antisera is diluted in PBS containing 5% guinea pig serum, as a complement source. *Chlamydiae* (10^4 IFU; infectious units) are added to the antisera dilutions. The antigen-antibody mixtures are incubated at 37EC for 45 minutes and inoculated into duplicate confluent Hep-2 or HeLa cell monolayers contained in glass vials (e.g., 15 by 45 mm), which have been washed twice with PBS prior to inoculation. The monolayer cells are infected by centrifugation at 1000X g for 1 hour followed by stationary incubation at 37E for 1 hour. Infected monolayers are incubated for 48 or 72 hours, fixed and stained with a *Chlamydiae* specific antibody, such as anti-MOMP for *C.trachomatis, etc.* IFUs are counted in ten fields at a magnification of 200X. Neutralization titer is assigned based on the dilution that gives 50% inhibition as compared to control monolayers/IFU.

The efficacy of vaccine compositions can be determined *in vivo* by challenging animal models of *Chlamydia pneumoniae* infection, e.g., mice or rabbits, with the vaccine compositions. For example, *in vivo* vaccine composition challenge studies can be performed in the murine model of *Chlamydia pneumonia* infection described by Moazed et al. (1997). Briefly, male homozygous apoE deficient and/or C57 BL/6J mice are immunized with vaccine compositions. Post-vaccination, the mice are mildly sedated by subcutaneous injection of a mixture of ketamine and xylazine, and inoculated intranasally with a total volume of 0.03-0.05 ml of organisms suspended in SPG medium or with SPG alone. The inoculations of *Chlamydia pneumoniae* are approximately 3×10^7 IFU/mouse. The mice are inoculated with *Chlamydia pneumoniae* at 8, 10, and 12 weeks of age. Tissues are then collected from the lung, spleen, heart, etc. at 1-20 weeks after the first inoculation. The presence of organisms is scored using PCR, histology and immunocytochemistry, or by quantitative culture/IFU after tissue homogenization.

Alternatively, *in vivo* vaccine composition challenge studies can be performed in the rabbit model of *Chlamydia pneumoniae* described by Laitinen et al. (1997). Briefly, New Zealand white rabbits (5 months old) are immunized with the vaccine compositions. Post-vaccination, the rabbits are sedated with Hypnorm, 0.3 ml/Kg of body weight, intramuscularly, and inoculated intranasally with a total of 0.5 ml of *Chlamydia pneumoniae* suspended in SPG medium or with SPG alone. The inoculations of *Chlamydia pneumoniae* are approximately $3x10^7$ IFU/rabbit. The rabbits are reinfected in the same manner and with the same dose 3 weeks after the primary inoculation. Tissues are then collected 2 weeks after the primary infection and 1, 2, and 4 weeks after the reinfection. The presence of *Chlamydia pneumoniae* is scored using PCR, histology and immunocytochemistry, or by quantitative culture/IFU after tissue homogenization.

The vaccine compositions comprising nucleotide sequences or vectors into which the said sequences are inserted are in particular described in International Application No. WO 90/11092 and also in International Application No. WO 95/11307.

The nucleotide sequence constituting the vaccine composition according to the invention may be injected into the host after having been coupled to compounds which promote the penetration of this polynucleotide inside the cell or its transport up to the cell nucleus. The resulting conjugates may be encapsulated into polymeric microparticles, as described in International Application No. WO 94/27238 (Medisorb Technologies International).

According to another embodiment of the vaccine composition according to the invention, the nucleotide sequence, preferably a DNA, is complexed with the DEAE-dextran (Pagano et al., 1967) or with nuclear proteins (Kaneda et al., 1989), with lipids (Felgner et al., 1987) or encapsulated into liposomes (Fraley et al., 1980) or alternatively introduced in the form of a gel facilitating its transfection into the cells (Midoux et al., 1993, Pastore et al., 1994). The polynucleotide or the vector according to the invention may also be in suspension in a buffer solution or may be combined with liposomes.

Advantageously, such a vaccine will be prepared in accordance with the technique described by Tacson et al. or Huygen et al. in 1996 or alternatively in accordance with the technique described by Davis et al. in International Application No. WO 95/11307.

Such a vaccine may also be prepared in the form of a composition containing a vector according to the invention, placed under the control of regulatory elements allowing its expression in humans or animals. It is possible, for example, to use, as vector for the *in vivo* expression of the polypeptide antigen of interest, the plasmid pcDNA3 or the plasmid pcDNA1/neo, both marketed by Invitrogen ® & D Systems, Abingdon, United Kingdom). It is also possible to use the plasmid V1Jns.tPA, described by Shiver et al. in 1995. Such a vaccine will advantageously comprise, in addition to the recombinant vector, a saline solution, for example a sodium chloride solution.

The immunogenic compositions of the invention can also be utilized as part of methods for immunization, wherein such methods comprise administering to a host, e.g., a human host, an

immunizing amount of the immunogenic compositions of the invention. In a preferred embodiment, the method of immunizing is a method of immunizing against *Chlamydia pneumoniae*.

A pharmaceutically acceptable vehicle is understood to designate a compound or a combination of compounds entering into a pharmaceutical or vaccine composition which does not cause side effects and which makes it possible, for example, to facilitate the administration of the active compound, to increase its life and/or its efficacy in the body, to increase its solubility in solution or alternatively to enhance its preservation. These pharmaceutically acceptable vehicles are well known and will be adapted by persons skilled in the art according to the nature and the mode of administration of the active compound chosen.

As regards the vaccine formulations, these may comprise appropriate immunity adjuvants which are known to persons skilled in the art, such as, for example, aluminum hydroxide, a representative of the family of muramyl peptides such as one of the peptide derivatives of N-acetyl-muramyl, a bacterial lysate, or alternatively incomplete Freund's adjuvant, Stimulon™ QS-21 (Aquila Biopharmaceuticals, Inc., Framingham, MA), MPL™ (3-O-deacylated monophosphoryl lipid A; RIBI ImmunoChem Research, Inc., Hamilton, MT), aluminum phosphate, IL-12 (Genetics Institute, Cambridge, MA).

Preferably, these compounds will be administered by the systemic route, in particular by the intravenous route, by the intranasal, intramuscular, intradermal or subcutaneous route, or by the oral route. More preferably, the vaccine composition comprising polypeptides according to the invention will be administered several times, spread out over time, by the intradermal or subcutaneous route.

Their optimum modes of administration, dosages and galenic forms may be determined according to criteria which are generally taken into account in establishing a treatment adapted to a patient, such as for example the patient's age or body weight, the seriousness of his general condition, tolerance of the treatment and the side effects observed.

The invention comprises the use of a composition according to the invention for the treatment or the prevention of cardiovascular diseases, preferably linked to the presence of atheroma, which are induced or worsened by *Chlamydia pneumoniae*.

Finally, the invention comprises the use of a composition according to the invention for the treatment or the prevention of respiratory diseases which are induced or worsened by the presence of *Chlamydia pneumoniae*, preferably asthma.

Other characteristics and advantages of the invention appear in the following examples and figures:

<u>Legend to the figures :</u>

Figure 1 : Line for the production of *Chlamydia pneumoniae* sequences

Figure 2 : Analysis of the sequences and assembling

Figure 3 : Finishing techniques

Figure 3a) : Assembly map

Figure 3b) : Determination and use of the orphan ends of the contigs

EXAMPLES

Experimental procedures

Cells

The *Chlamydia pneumoniae* strain (CM1) used by the inventors is obtained from ATCC (American Culture Type Collection) where it has the reference number ATCC 1360-VR.

It is cultured on HeLa 229 cells, obtained from the American Type Culture Collection, under the reference ATCC CCL-2.1.

Culture of the cells

The HeLa ATCC CCL-2.1 cells are cultured in 75-ml cell culture flasks (Corning). The culture medium is Dulbecco's modified cell culture medium (Gibco BRL No. 04101965) supplemented with MEM amino acids (Gibco BRL - No. 04301140) L (5 ml per 500 ml of medium) and 5% foetal calf serum (Gibco BRL No. 10270 batch 40G8260K) without antibiotics or antifungals.

The cell culture stock is maintained in the following manner. The cell cultures are examined under an inverted microscope. 24 hours after confluence, each cellular lawn is washed with PBS (Gibco BRL No. 04114190), rinsed and then placed for 5 min in an oven in the presence of 3 ml of trypsine (Gibco BRL No. 25200056). The cellular lawn is then detached and then resuspended in 120 ml of culture medium, the whole is stirred in order to make the cellular suspension homogeneous. 30 ml of this suspension are then distributed per cell culture flask. The flasks are kept in a CO_2 oven (5%) for 48 hours at a temperature of 37°C. The cell stock is maintained so as to have available daily 16 flasks of subconfluent cells. It is these subconfluent cells which will be used so as to be infected with Chlamydia. 25-ml cell culture flasks are also used, these flasks are prepared in a similar manner but the volumes used for maintaining the cells are the following: 1 ml of trypsine, 28 ml of culture medium to resuspend the cells, 7 ml of culture medium are used per 25-ml flask.

Infection of the cells with Chlamydia

Initially, the Chlamydiae are obtained frozen from ATCC (-70°C), in suspension in a volume of 1 ml. This preparation is slowly thawed, 500 µl are collected and brought into contact with subconfluent cells, which are obtained as indicated above, in a 25-ml cell culture flask, containing 1 ml of medium, so as to cover the cells. The flask is then centrifuged at 2000 rpm in a "swing" rotor for microtitre plates, the centrifuge being maintained at a temperature of 35°C. After centrifugation,

the two flasks are placed in an oven at 35°C for three hours. 6 ml of culture medium containing cycloheximide (1 µg/ml) are then added and the flask is stored at 35°C. After 72 hours, the level of infection is evaluated by direct immunofluorescence and by the cytopathogenic effect caused to the cells.

Direct immunofluorescence

Starting with infected cells, which were obtained as indicated above, a cellular smear is deposited with a Pasteur pipette on a microscope slide. The cellular smear is fixed with acetone for 10 minutes; after draining the acetone, the smear is covered with 30 µl of murine monoclonal antibodies directed against MOMP (major outer membrane protein) of Chlamydia (Syva, Biomérieux) labelled with fluorescein isothiocyanate. The whole is then incubated in a humid chamber at a temperature of 37°C. The slides are then rinsed with water, slightly dried, and then after depositing a drop of mounting medium, a coverslip is mounted before reading. The reading is carried out with the aid of a fluorescence microscope equipped with the required filters (excitation at 490 nm, emission at 520 nm).

Harvesting of the *Chlamydia pneumoniae*

After checking the infection by direct immunofluorescence, carried out as indicated above, the culture flasks are opened under a sterile cabinet, sterile glass beads with a diameter of the order of a millimeter are placed in the flask. The flask is closed and then vigorously stirred while being maintained horizontally, the cellular lawn at the bottom, so that the glass beads can have a mechanical action on the cellular lawn. Most of the cells are thus detached or broken; the effect of the stirring is observed under an optical microscope so as to ensure proper release of Chlamydiae.

Large-scale infection of the cell cultures

The product of the Chlamydiae harvest (culture medium and cellular debris) is collected with a pipette, and distributed into three cell culture flasks containing subconfluent HeLa ATCC CCL-2.1 cells, obtained as indicated above. The cells thus inoculated are placed under gentle stirring (swing) in an oven at 35°C. After one hour, the flasks are kept horizontally in an oven so that the culture medium covers the cells for 3 hours. 30 ml of culture medium containing actydione (1 µg/ml) are then added to each of the flasks. The culture flasks are then stored at 35°C for 72 hours. The cells thus infected are examined under an optical microscope after 24 hours, the cytopathogenic effect is evaluated by the appearance of cytoplasmic inclusions which are visible under an inverted optical microscope. After 72 hours, the vacuoles containing the Chlamydiae occupy the cytoplasm of the cell and push the cell nucleus sideways. At this stage, numerous cells are spontaneously destroyed and have left free elementary bodies in the culture medium. The Chlamydiae are harvested as described above and are either frozen at –80°C or used for another propagation.

Purification of the Chlamydiae

The product of the Chlamydia harvests is stored at -80°C and thawed on a water bath at

room temperature. After thawing, each tube is vigorously stirred for one minute and immersed for one minute in an ultrasound tank (BRANSON 1200); the tubes are then stirred by inverting before being centrifuged for 5 min at 2000 rpm. The supernatant is carefully removed and kept at cold temperature (ice). The supernatant is vigorously stirred and then filtered on nylon filters having pores of 5 microns in diameter on a support (Nalgene) allowing a delicate vacuum to be established under the nylon filter. For each filtration, three nylon filters are superposed; these filters are replaced after every 40 ml of filtrate. Two hundred milliliters of filtration product are kept at cold temperature, and then after stirring by inverting, are centrifuged at 10,000 rpm for 90 min, the supernatant is removed and the pellet is taken up in 10 ml of 10 mM Tris, vigorously vortexed and then centrifuged at 10,000 rpm for 90 min. The supernatant is removed and the pellet is taken up in a buffer (20 mM Tris pH 8.0, 50 mM KCl, 5 mM $MgCl_2$) to which 800 units of DNAse I (Boehringer) are added. The whole is kept at 37°C for one hour. One ml of 0.5 M EDTA is then added, the whole is vortexed and frozen at –20°C.

Preparation of the DNA

The Chlamydiae purified above are thawed and subjected to a proteinase K (Boehringer) digestion in a final volume of 10 ml. The digestion conditions are the following: 0.1 mg/ml proteinase K, 0.1 × SDS at 55EC, stirring every 10 min. The product of digestion is then subjected to a double extraction with phenol-chloroform, two volumes of ethanol are added and the DNA is directly recovered with a Pasteur pipette having one end in the form of a hook. The DNA is dried on the edge of the tube and then resuspended in 500 µl of 2 mM Tris pH 7.5. The DNA is stored at 4°C for at least 24 hours before being used for the cloning.

Cloning of the DNA

After precipitation, the DNA is quantified by measuring the optical density at 260 nm. Thirty µg of Chlamydia DNA are distributed into 10 tubes of 1.5 ml and diluted in 300 µl of water. Each of the tubes is subjected to 10 applications of ultrasound lasting for 0.5 sec in a sonicator (unisonix XL2020). The contents of the 10 tubes are then grouped and concentrated by successive extractions with butanol (Sigma B1888) in the following manner: two volumes of butanol are added to the dilute DNA mixture. After stirring, the whole is centrifuged for five minutes at 2500 rpm and the butanol is removed. This operation is repeated until the volume of the aqueous phase is less than 1 ml. The DNA is then precipitated in the presence of ethanol and of 0.5 M sodium acetate pH 5.4, and then centrifuged for thirty minutes at 15,000 rpm at cold temperature (4°C). The pellet is washed with 75% ethanol, centrifuged for five minutes at 15,000 rpm and dried at room temperature. A tenth of the preparation is analysed on a 0.8% agarose gel. Typically, the size of the DNA fragments thus prepared is between 200 and 8000 base pairs.

To allow the cloning of the DNA obtained, the ends are repaired. The DNA is distributed in an amount of 10 µg/tube, in the following reaction medium: 100 µl final volume, 1 × buffer

(Biolabs 201L), 0.5 µl BSA 0.05 mg/ml, 0.1 mM dATP, 0.1 mM each of dGTP, dCTP or dTTP, 60,000 IU T4 DNA polymerase. The reaction is incubated for thirty minutes at 16°C. The contents of each of the tubes are then grouped before carrying out an extraction with phenol-chloroform and then precipitating the aqueous phase as described above. After this step, the DNA thus prepared is phosphorylated. For that, the DNA is distributed into tubes in an amount of 10 µg per tube, and then in a final volume of 50 µl, the reaction is prepared in the following manner: 1 mM ATP, 1 × kinase buffer, 10 IU T4 polynucleotide kinase (Biolabs 201L). The preparation is incubated for thirty minutes at 37°C. The contents of the tubes are combined and a phenol-chloroform extraction and then a precipitation are carried out in order to precipitate the DNA. The latter is then suspended in 1 µl of water and then the DNA fragments are separated according to their size on a 0.8% agarose gel (1 × TAE). The DNA is subjected to an electric field of 5 V/cm and then visualized on a UV table. The fragments whose size varies between 1200 and 2000 base pairs are selected by cutting out the gel. The gel fragment thus isolated is placed in a tube and then the DNA is purified with the Qiaex kit (20021 Qiagen), according to the procedure provided by the manufacturer.

<u>Preparation of the vector</u>

14 µg of the cloning vector pGEM-5Zf (Proméga P2241) are diluted in a final volume of 150 µl and are subjected to digestion with the restriction enzyme EcoRV 300 IU (Biolabs 195S) according to the protocol and with the reagents provided by the manufacturer. The whole is placed at 37°C for 150 min and then distributed in the wells of a 0.8% agarose gel subjected to an electric field of 5 V/cm. The linearized vector is visualized on a UV table, isolated by cutting out the gel and then purified by the Qiaex kit (Qiagen 20021) according to the manufacturer's recommendations. The purification products are grouped in a tube, the volume is measured and then half the volume of phenol is added and the whole is vigorously stirred for 1 min. Half the volume of chloroform-isoamyl alcohol 24:1 is added and vigorously stirred for 1 min. The whole is centrifuged at 15,000 rpm for 5 min at 4°C, the aqueous phase is recovered and transferred into a tube. The DNA is precipitated in the presence of 0.3 M sodium acetate, pH 5.4 and 3 volumes of ethanol and placed at -20°C for 1 hour. The DNA is then centrifuged at 15,000 rpm for 30 min at 4°C, the supernatant is removed while preserving the pellet, washed twice with 70% ethanol. After drying at room temperature, the DNA is suspended in 25 µl of water.

<u>Phosphorylation of the vector</u>

25 µl of the vector prepared in the preceding step are diluted in a final volume of 500 µl of the following reaction mixture:

After repair, the DNA is subjected to a phenol-chloroform extraction and a precipitation, the pellet is then taken up in 10 µl of water, the DNA is quantified by measuring the optical density at 260 nm. The quantified DNA is ligated into the vector PGEm-5Zf(+) prepared by the restriction

enzyme EcoRV and dephosphorylated (see preparation of the vector). The ligation is carried out under three conditions which vary in the ratio between the number of vector molecules and the number of insert molecules. Typically, an equimolar ratio, a ratio of 1:3 and a ratio of 3:1 are used for the ligations which are, moreover, carried out under the following conditions: vector PGEm-5Zf(+) 25 ng, cut DNA, ligation buffer in a final volume of 20 µl with T4 DNA ligase (Amersham E70042X); the whole is then placed in a refrigerator overnight and then a phenol-chloroform extraction and a precipitation are carried out in a conventional manner. The pellet is taken up in 5 µl of water.

Transformation of the bacteria

Plating of the bacteria

Petri dishes containing LB Agar medium containing ampicillin (50 µg/ml), Xgal (280 µg/ml) [5-bromo-4-chloro-indolyl-beta-D-galactopyranoside (Sigma B-4252)], IPTG (140 µg/ml) [isopropyl-beta-D-thiogalactoside (Sigma I-6758)] are used, 50 and 100 µl of bacteria are plated for each of the ligations. The Petri dishes are placed upside down at 37°C for 15 to 16 hours in an oven. The number of "recombinant" positive clones is evaluated by counting the white colonies and the blue colonies which are thought to contain the vector alone.

Evaluation of the "recombinant" positive clones

Ninety-four white colonies and two blue colonies are collected with the aid of sterile cones and are deposited at the bottom of the wells of plates designed for carrying out the amplification techniques. 30 µl of the following reaction mixture are added to each well: 1.7 mM $MgCl_2$, 0.2 mM each of dATP, dCTP, dGTP and dTTP, two synthetic oligonucleotides corresponding to sequences flanking the cloning site on either side and orienting the synthesis of the DNA in a convergent manner (0.5 µM RP and PU primers, 1 U TAQ polymerase (GibcoBRL 18038-026)).

The colonies thus prepared are subjected to a temperature of 94°C for 5 min and then to 30 thermal cycles composed of the following steps: 94°C for 40 s, 50°C for 30 s, 72°C for 180 s. The reaction is then kept for 7 min at 72°C and then kept at 4°C.

The amplification products are deposited on an agarose gel (0.8%), stained with ethidium bromide, subjected to electrophoresis, and then analysed on an ultraviolet table. The presence of an amplification fragment having a size greater than 500 base pairs indicates the presence of an insert. The bacterial clones are then prepared so as to study the sequence of their insert.

Sequencing

To sequence the inserts of the clones obtained as above, these were amplified by PCR on bacteria cultures carried out overnight using the primers for the vectors flanking the inserts. The sequence of the ends of these inserts (on average 500 bases on each side) was determined by automated fluorescent sequencing on an ABI 377 sequencer, equipped with the ABI Prism DNA Sequencing Analysis software (version 2.1.2).

Analysis of the sequences

The sequences obtained by sequencing in a high-yield line (Figure 1) are stored in a database; this part of the production is independent of any treatment of the sequences. The sequences are extracted from the database, avoiding all the regions of inadequate quality, that is to say the regions for which uncertainties are observed on the sequence at more than 95%. After extraction, the sequences are introduced into a processing line, the diagram of which is described in Figure 2. In a first path of this processing line, the sequences are assembled by the Gap4 software from R. Staden (Bonfield et al., 1995) (OS UNIX/SUN Solaris); the results obtained by this software are kept in the form of two files which will be used for a subsequent processing. The first of these files provides information on the sequence of each of the contigs obtained. The second file represents all the clones participating in the composition of all the contigs as well as their positions on the respective contigs.

The second processing path uses a sequence assembler (TIGR-Asmg assembler UNIX/SUN Solaris); the results of this second processing path are kept in the form of a file in the TIGR-Asmg format which provides information on the relationship existing between the sequences selected for the assembly. This assembler is sometimes incapable of linking contigs whose ends overlap over several hundreds of base pairs.

The results obtained from these two assemblers are compared with the aid of the BLAST program, each of the contigs derived from one assembly path being compared with the contigs derived from the other path.

For the two processing paths, the strict assembly parameters are fixed (95% homology, 30 superposition nucleotides). These parameters avoid 3 to 5% of the clones derived from eukaryotic cells being confused with sequences obtained from the clones derived from *Chlamydia pneumoniae*. The eukaryotic sequences are however preserved during the course of this project; the strategy introduced, which is described below, will be designed, inter alia, not to be impeded by these sequences derived from contaminating clones.

The results of these two assemblers are processed in a software developed for this project. This software operates on a Windows NT platform and receives, as data, the results derived from the STADEN software and/or the results derived from the TIGR-Asmg assembler, the software, results, after processing of the data, in the determination of an assembly map which gives the proximity relationship and the orientation of the contigs in relation to one another (Figure 3a). Using this assembly map, the software determines all the primers necessary for finishing the project. This treatment, which will be detailed below, has the advantage of distinguishing the isolated sequences derived from the contaminations, by the DNA eukaryotic cells, of the small-sized sequences clearly integrated into the project by the relationships which they establish with contigs. In order to allow, without any risk of error, the arrangement and the orientation of the contigs in relation to one another, a statistical evaluation of the accuracy of the names (naming) “naming” of sequence is made from the results of “contigation”. This evaluation makes it possible to give each of the clone plates, as well as each of the subsets of plates, a weight which is inversely proportional to probable error rate existing in

the "naming" of the sequences obtained from this plate or from a subset of this plate. In spite of a low error rate, errors may occur throughout the steps of production of the clones and of the sequences. These steps are numerous, repetitive and although most of them are automated, others, like the deposition in the sequencers, are manual; it is then possible for the operator to make mistakes such as the inversion of two sequences. This type of error has a repercussion on the subsequent processing of the data, by resulting in relationships (between the contigs) which do not exist in reality, then in attempts at directed sequencing between the contigs which will end in failure. It is because of this that the evaluation of the naming errors is of particular importance since it allows the establishment of a probabilistic assembly map from which it becomes possible to determine all the clones which will serve as template to obtain sequences separating two adjacent contigs. Table 2 of parent U.S. application serial No. 60/107078 filed November 4, 1998 and French application 97-14673 filed November 21, 1997, each of which is incorporated by reference herein in its entirety, gives the clones and the sequences of the primers initially used during the initial operations.

To avoid the step which consists in ordering and then preparing the clones by conventional microbiological means, outer and inner primers oriented towards the regions not yet sequenced are defined by the software. The primers thus determined make it possible to prepare, by PCR, a template covering the nonsequenced region. It is the so-called outer primers (the ones most distant from the region to be sequenced) which are used to prepare this template. The template is then purified and a sequence is obtained on each of the two strands during 2 sequencing reactions which each use one of the 2 inner primers. In order to facilitate the use of this approach, the two outer primers and the two inner primers are prepared and then stored on the same position of 4 different 96-well plates. The two plates containing the outer primers are used to perform the PCRs which will serve to prepare the templates. These templates will be purified on purification columns preserving the topography of the plates. Each of the sequences will be obtained using primers situated on one and then on the other of the plates containing the inner primers. This distribution allows a very extensive automation of the process and results in a method which is simple to use for finishing the regions not yet sequenced. Table 3 of parent U.S. application serial No. 60/107078 filed November 4, 1998 and French application 97-14673 filed November 21, 1997, each of which is incorporated by reference herein in its entirety, gives the names and the sequences of the primers used for finishing *Chlamydia pneumoniae*.

Finally, a number of contigs exist in a configuration where one of their ends is not linked to any other contig end (Figure 3b) by a connecting clone relationship (a connecting clone is defined as a clone having one sequence end on a contig and the other end of its sequence on another contig; furthermore, this clone must be derived from a plate or a subset of plates with adequate naming quality). For the *Chlamydia pneumoniae* project, this particular case occurred 24 times. Two adjacent PCR primers orienting the synthesis of the DNA towards the end of the consensus sequence are defined for each of the orphan ends of the consensus sequence. The primer which is closest to the end

of the sequence is called the inner primer whereas the primer which is more distant from the end of the sequence is called the outer primer. The outer primers are used to explore the mutual relationship between the orphan ends of the different contigs. The presence of a single PCR product and the possibility of amplifying this product unambiguously using the inner primers evokes the probable relationship between the contigs on which the primers which allowed the amplification are situated. This relationship will be confirmed by sequencing and will allow the connection between the orphan ends of the consensus sequences. This strategy has made it possible to obtain a complete map of the *Chlamydia pneumoniae* chromosome and then to finish the project.

Quality control

All the bases not determined with certainty in the chromosomal sequence were noted and the density of uncertainties was measured on the entire chromosome. The regions with a high density of uncertainties were noted and the PCR primers spanning these regions were drawn and are represented in Table 4 of parent U.S. application serial No. 60/107078 filed November 4, 1998 and French application 97-14673 filed November 21, 1997 each of which is incorporated by reference herein in its entirety.

The sequence of each of the PCR products was obtained with two operational primers different from the amplification primers. The sequences were obtained in both directions for all the PCRs (100% success).

Data banks

Local reorganizations of major public banks were used. The protein bank used consists of the nonredundant fusion of the Genpept bank (automated translation of GenBank, NCBI; Benson et al., 1996).

The entire BLAST software (public domain, Altschul et al., 1990) for searching for homologies between a sequence and protein or nucleic data banks was used. The significance levels used depend on the length and the complexity of the region tested as well as the size of the reference bank. They were adjusted and adapted to each analysis.

The results of the search for homologies between a sequence according to the invention and protein or nucleic data banks are presented and summarized in Table 1 below.

Table 1: List of coding chromosome regions and homologies between these regions and the sequence banks.

Legend to Table 1: Open reading frames are identified with the GenMark software version 2.3A (GenePro), the template used is *Chlamydia pneumoniae* of order 4 on a length of 196 nucleotides with a window of 12 nucleotides and a minimum signal of 0.5. The reading frames ORF2 to ORF 1137 are numbered in order of appearance on the chromosome, starting with ORF2 (ORF column). The positions of the beginning and of the end are then given in column 2 (position). When the position of the beginning is greater than the position of the end, this means that the region is

encoded by the strand complementary to the sequence which was given in the sequence SEQ ID No. 1.

All the putative products were subjected to a search for homology on GENPEPT (release 102 for SEQ ID No. 2 to SEQ ID No. 1137, and release 108 for SEQ ID No. 1138 to SEQ ID No. 1291 and SEQ ID No. 6844 to SEQ ID No. 6849) with the BLASTP software (Altschul et al. 1990). With, as parameters, the default parameters with the exception of the expected value E set at 10^{-5} (for SEQ ID No. 2 to SEQ ID No. 1137) and P value set at e^{-10} (for SEQ ID No. 1138 to SEQ ID No. 1291 and SEQ ID No. 6844 to SEQ ID No. 6849). Subsequently, only the identities greater than 30% (I% column) were taken into account. The description of the most homologous sequence is given in the Homology column; the identifier for the latter sequence is given in the ID column and the animal species to which this sequence belongs is given in the Species column. The Homology score is evaluated by the sum of the blast scores for each region of homology and reported in the Score column.

Materials and Methods for transmembrane domains:

The DAS software was used as recommended by the authors (Cserzo et al., 1997).

This method uses, to predict the transmembrane domains, templates derived from a sampling of selected proteins. All the regions for which a "Cutoff" greater than 1.5 was found by the program were taken into account.

Additional ORF Finder Programs

For this analysis, two additional ORF finder programs were used to predict potential open reading frames of a minimum length of 74 amino acids; Glimmer (Salzberg, S.L., Delcher, A., Kasif, S., and W. White. 1998. Microbial gene identification using interpolated Markov models. Nucleic Acids Res. 26:544-548.), and an in-house written program. The in-house program used a very simple search algorithm. The analysis required the that the genomic DNA sequence text be in the 5' to 3' direction, the genome is circular, and that TAA, TAG, and TGA are stop codons. The search parameters were as follows:

(1) A search for an ORF that started with a GTG codon was performed. If no GTG codons were found, then a search for an ATG codon was performed. However, if a GTG codon was found, then a search downstream for a ATG codon was performed. All start and stop nucleotide positions were recorded.

(2) A search for an ORF that started with a TTG codon was performed. If no TTG codons were found, then a search for a ATG codon was performed. However, if a TTG codon was found, then a search downstream for a ATG codon was performed. All start and stop nucleotide positions were recorded.

(3) The analysis described in steps 1 and 2 were repeated for the opposite strand of DNA sequence.

(4) A search for ORFs that determined all ORF lengths using start and stop positions in the same reading frames was performed.

(5) All ORFs whose DNA length was less than 225 nucleotides were eliminated from the search.

Surface Exposed Protein Search Criteria

Potential cell surface vaccine targets are outer membrane proteins such as porins, lipoproteins, adhesions and other non-integral proteins. In *Chlamydia psittaci*, the major immunogens is a group of putative outer membrane proteins (POMPs) and no homologs have been found in *Chlamydia pneumoniae* and *Chlamydia trachomatis* by traditional analysis (Longbottom, D., Russell, M., Dunbar, S.M., Jones, G.E., and A.J. Herring. 1998. Molecular Cloning and Characterization of the Genes Coding for the Highly Immunogenic Cluster of 90-Kilodalton Envelope Proteins from *Chlamydia psittaci* Subtype That Causes Abortion in Sheep. Infect Immun 66:1317-1324.) Several putative outer membrane proteins have been identified in *Chlamydia pneumoniae*, all of which may represent vaccine candidates. The major outer membrane protein (MOMP) gene (omp1) has been found in various isolates of *Chlamydia pneumoniae* (Jantos, CA., Heck, S., Roggendorf, R., Sen-Gupta, M., and Hegemann, JH. 1997. Antigenic and molecular analyses of different chlamydia pneumoniae strains. J. Clin Microbiology 35(3):620-623.) Various criteria, as listed below, were used to identify putative surface exposed ORFs from the genomic DNA sequence of *Chlamydia pneumoniae* (French application 97-14673 filed 21 November 1997). Any ORF which met any one or more of the individual criteria were listed in this category.

Protein homology searches were done using the Blastp 2.0 tool (Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.) An ORF product was labeled surface exposed if there was homology to a known, or hypothetical, or putative surface exposed protein with a P score better than e^{-10}.

Most, if not all, proteins that are localized to the membrane of bacteria, via a secretory pathway, contain a signal peptide. A software program, SignalP, analyzes the amino acid sequence of an ORF for such a signal peptide (Nielsen, H., Engelbrecht. J., Brunak, S., and G. von Heijne. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Engineering 10:1-6.) The first 60 N-terminal amino acids of each ORF were analyzed by SignalP using the Gram-Negative software database. The output generates four separate values, maximum C, maximum Y, maximum S, and mean S. The S-score, or signal region, is the probability of the position belonging to the signal peptide. The C-score, or cleavage site, is the probability of the position being the first in the mature protein. The Y-score is the geometric average of the C-score and a smoothed derivative of the S-score. A conclusion of either a Yes or No is given next to each score. If all four conclusions are Yes and the C-terminal amino acid is either a phenylalanine (F) or a tyrosine (Y), the ORF product was labelled outer membrane (Struyve, M., Moons, M., and J. Tommassen.

1991. Carboxy-terminal Phenylalanine is Essential for the Correct Assembly of a Bacterial Outer Membrane Protein. J. Mol. Biol. 218:141-148.)

The program called Psort, determines the localization of a protein based on its signal sequence, recognition of transmembrane segments, and analysis of its amino acid composition (Nakai, K., and M. Kanehisa. 1991. Expert system for predicting protein localization sites in gram-negative bacteria. Proteins 11:95-110.) An ORF product is considered to be an outer membrane protein if the output data predicts the protein as outer membrane with a certainty value of 0.5 or better and whose value is at least twice as large as the next predicted localized certainty value.

Finally, ORF products that were not predicted to be outer membrane or surface exposed, based on the above criteria, were further analyzed. The blastp output data for these ORFs were searched using various general and specific keywords, suggestive of known cell surface exposed proteins. An ORF was labeled surface exposed if the keywords matched had a Blastp hit, had a P score better than e^{-10}, and that there was no better data indicating otherwise. The following is a list of the searched keywords:

Adhesion	Adhesin	Invasin	Invasion	Extensin	
Omp	Outer Surface	Porin	Outer Membrane		
Cell Surface	Cell Wall	Pilus	Pilin	Flagellar sheath	BtuB
Cir	ChuA	CopB	ExeD	FadL	FecA
FepA	FhuA	FmdC	FomA	FrpB	GspD
HemR	HgbA	Hgp	HmbR	HmuR	HMW
HrcC	Hrp	InvG	LamB	LbpA	LcrQ
Lmp1	MxiD	MOMP	PilE	HpaA	NolW
NspA	OpcP	OpnP	Opr	OspA	PhoE
PldA	Por	PscC	PulD	PupA	QuiX
RafY	ScrY	SepC	ShuA	SomA	SpiA
Tbp1	Yop	YscC	mip	Tol	

Those ORFs that did not meet the minimum requirement for being an outer membrane protein based on the above search criteria but which were homologous to identified outer membrane ORFs in *Chlamydia trachomatis* were included. The *Chlamydia trachomatis* genome (French patent applications FR97-15041, filed 28 November 1997 and 97-16034 filed 17 December 1997) was analyzed using the above search criteria and a number of outer membrane ORFs were identified. These *Chlamydia trachomatis* ORFs were then tested against the *Chlamydia pneumoniae* genome using Blastp. Any *Chlamydia pneumoniae* ORF with a Blastp P value better than e^{-10} against a *Chlamydia trachomatis* outer membrane was included in this section, if there was no better data

indicating otherwise. A list of ORFs in the *Chlamydia pneumoniae* genome encoding putative surface exposed proteins is set forth above in the specification.

Identification of Putative Lipoproteins in the Genome of *Chlamydia pneumoniae*

Lipoproteins are the most abundant post-translationally modified bacterial secretory proteins (Pugsley, A. P.. 1993. The complete general secretory pathway in Gram-negative bacteria. Microbiol. Rev. 57:50-108). The characteristic features of lipoproteins are a thiol-linked diacylglyceride and an amine-linked monoacyl group on the cysteine that becomes the amino-terminal residue after signal peptide cleavage by Signal Peptidase II. (Pugsley, A. P.. 1993. The complete general secretory pathway in Gram-negative bacteria. Microbiol. Rev. 57:50-108). The identification of putative lipoproteins from the genomic sequencing of *Chlamydia pneumoniae* was done by examining the deduced amino acid sequence of identified ORFs for the presence of a signal peptide with a Signal Peptidase II cleavage site analogous to the consensus sequence for prolipoprotein modification and processing reactions (Hayashi, S., and H. C. Wu. 1992. Identification and characterization of lipid-modified proteins in bacteria, p. 261-285. *In* N. M. Hooper and A. J. Turner (ed.) Lipid modification of proteins: A practical approach. Oxford University Press, New York; Sutcliffe, I. C. and R. R. B. Russell. 1995. Lipoproteins of Gram-positive bacteria. J. Bacteriol. 177:1123-1128.).

Chlamydia pneumoniae ORFs were initially screened for the most basic of lipoprotein characteristics, a cysteine in the first 30 amino acids of the deduced protein. ORFs with a standard start codon (ATG, GTG, or TTG) and having one or more of the following characteristics were selected for direct analysis of their first 30 amino acids:

(a) Significant Signal P value (at least two out of the four values are Yes)

(b) PSORT value indicating membrane passage (IM-inner membrane, Peri-periplasm, or OM-outer membrane)

(c) Identification of the word lipoprotein among the ORF blastp data set.

(d) A Blastp value of $<e^{-10}$ with a putative lipoprotein from *Chlamydia trachomatis* (French applications 97-15041 filed 28 November 1997 and 97-16034 filed 17 December 1997).

The first 30 amino acids of each ORF in this set were analyzed for the characteristics commonly found in lipoprotein signal peptides (Pugsley, A. P.. 1993. The complete general secretory pathway in Gram-negative bacteria. Microbiol. Rev. 57:50-108; Hayashi, S., and H. C. Wu. 1992.

Identification and characterization of lipid- modified proteins in bacteria, p. 261-285. *In* N. M. Hooper and A. J. Turner (ed.) Lipid modification of proteins: A practical approach. Oxford University Press, New York; Sutcliffe, I. C. and R. R. B. Russell. 1995. Lipoproteins of Gram-positive bacteria. J. Bacteriol. 177:1123-1128.) Putative lipoprotein signal peptides were required to have a cysteine between amino acid 10 and 30 and reach a minimum score of three based on the following criteria for lipoprotein signal peptides:

(a) Identification of specific amino acids in specific positions around the cysteine which are part of the consensus Signal Peptidase II cleavage site (Hayashi, S., and H. C. Wu. 1992. Identification and characterization of lipid-modified proteins in bacteria, p. 261-285. *In* N. M. Hooper and A. J. Turner (ed.) Lipid modification of proteins: A practical approach. Oxford University Press, New York); Sutcliffe, I. C. and R. R. B. Russell. 1995. Lipoproteins of Gram-positive bacteria. J. Bacteriol. 177:1123-1128). Since the identification of the cleavage site is the most important factor in identifying putative lipoproteins, each correctly positioned amino acid contributed toward reaching the minimum score of three. (b) A hydrophobic region rich in alanine and leucine prior to the cleavage site (Pugsley, A. P.. 1993. The complete general secretory pathway in Gram-negative bacteria. Microbiol. Rev. 57:50-108) contributed toward reaching the minimum score of three.

(c) A short stretch of hydrophilic amino acids greater than or equal to 1 usually lysine or arginine following the N-terminal methionine (Pugsley, A. P.. 1993. The complete general secretory pathway in Gram-negative bacteria. Microbiol. Rev. 57:50-108) contributed toward reaching the minimum score of three.

A list of ORFs in the *Chlamydia pneumoniae* genome encoding putative lipoproteins is set forth above in the specification.

<u>LPS-Related ORFs of *Chlamydia pneumoniae*</u>

Lipopolysaccharide (LPS) is an important major surface antigen of Chlamydia cells. Monoclonal antibodies (Mab) directed against LPS of *Chlamydia pneumoniae* have been identified that can neutralize the infectivity of *Chlamydia pneumoniae* both in vitro and in vivo (Peterson, E.M., de la Maza, L.M., Brade, L., Brade, H. 1998. Characterization of a Neutralizing Monoclonal Antibody Directed at the Lipopolysaccharide of *Chlamydia pneumonia*. Infect. Immun. Aug. 66(8):3848-3855.) Chlamydial LPS is composed of lipid A and a core oligosaccharide portion and is phenotypically of the rough type (R-LPS) (Lukacova, M., Baumann, M., Brade, L., Mamat, U., Brade, H. 1994. Lipopolysaccharide Smooth-Rough Phase Variation in Bacteria of the Genus *Chlamydia*. Infect. Immun. June 62(6):2270-2276.) The lipid A component is composed of fatty acids which serve to anchor LPS in the outer membrane. The core component contains sugars and sugar derivatives such as a trisaccharide of 3-deoxy-D-manno-octulosonic acid (KDO) (Reeves, P.R., Hobbs, M., Valvano, M.A., Skurnik, M., Whitfield, C., Coplin, D., Kido, N., Klena, J., Maskell, D.,

Raetz, C.R.H., Rick, P.D. 1996. *Bacterial Polysaccharide Synthesis and Gene Nomenclature* pp. 10071-10078, Elsevier Science Ltd.). The KDO gene product is a multifunctional glycosyltransferase and represents a shared epitope among the Chlamydia. For a review of LPS biosynthesis see, e.g., Schnaitman, C.A., Klena, J.D. 1993. Genetics of Lipopolysaccharide Biosynthesis in Enteric Bacteria. Microbiol. Rev. 57:655-682.

A text search of the ORF blastp results identified several genes that are involved in Chlamydial LPS production with a P score better than e^{-10}. The following key-terms were used in the text search: KDO, CPS (Capsular Polysaccharide Biosynthesis), capsule, LPS, rfa, rfb, rfc, rfe, rha, rhl, core, epimerase, isomerase, transferase, pyrophosphorylase, phosphatase, aldolase, heptose, manno, glucose, lpxB, fibronectin, fibrinogen, fucosyltransferase , lic, lgt, pgm, tolC, rol, ChoP, phosphorylcholine, waaF, PGL-Tb1. A list of ORFs in the *Chlamydia pneumoniae* genome encoding putative polypeptides involved in LPS biosynthesis is set forth above in the specification.

Type III And Other Secreted Products

Type III secretion enables gram-negative bacteria to secrete and inject pathogenicity proteins into the cytosol of eukaryotic host cells (Hueck, C. J., 1998. Type III Protein Secretion Systems in Bacterial Pathogens of Animals and Plants. In Microbiology and Molecular Biology Reviews. 62:379-433.) These secreted factors often resemble eukaryotic signal transduction factors, thus enabling the bacterium to redirect host cell functions (Lee, C.A., 1997. Type III secretion systems: machines to deliver bacterial proteins into eukaryotic cells? Trends Microbiol. 5:148-156.) In an attempt to corrupt normal cellular functions, Chlamydial pathogenicity factors injected into the host cytosol will nonetheless, as cytoplasmic constituents be processed and presented in the context of the Major Histocompatibility Complex (MHC class I). As such, these pathogenicity proteins represent MHC class I antigens and will play an important role in cellular immunity. Also included in this set are secreted non-type III products that may play a role as vaccine components.

A text search of the ORF blastp results identified genes that are involved in *Chlamydia pneumoniae* protein secretion with a P score better than e^{-10}. The following key-terms were used in the text search in an effort to identify surface localized or secreted products: Yop, Lcr, Ypk, Exo, Pcr, Pop, Ipa, Vir, Ssp, Spt, Esp, Tir, Hrp, Mxi, hemolysin, toxin, IgA protease, cytolysin, tox, hap, secreted and Mip.

Chlamydia pneumoniae ORFs that did not meet the above keyword search criteria, but have homologs *in Chlamydia trachomatis* that do meet the search criteria are included herein. The *Chlamydia trachomatis* genome (French patent applications FR97-15041, filed 28 November 1997 and 97-16034 filed 17 December 1997) was analyzed using the above search criteria and a number of ORFs were identified. These *Chlamydia trachomatis* ORFs were tested against the *Chlamydia pneumoniae* genome using Blastp. Any *Chlamydia pneumoniae* ORF with a Blastp P value < e^{-10} against a *Chlamydia trachomatis* homolog, identified using the above search criteria, was included. A

list of ORFs in the *Chlamydia pneumoniae* genome encoding putative secreted proteins is in the specification.

Chlamydia pneumoniae: RGD Recognition Sequence

Proteins that contain Arg-Gly-Asp (RGD) attachment site, together with integrins that serve as their receptor constitute a major recognition system for cell adhesion. The RGD sequence is the cell attachment site of a large number of adhesive extracellular matrix, blood, and cell surface proteins and nearly half of the known integrins recognize this sequence in their adhesion protein ligands. There are many RGD containing microbial proteins such as the penton protein of adenovirus, the coxsackie virus, the foot and mouth virus and pertactin, a 69 kDa (kilodalton) surface protein of *Bordetella pertussis,* that serve as ligands through which these microbes bind to integrins on the cell surfaces and gain entry into the cell. The following provides evidence supporting the importance of RGD in microbial adhesion:

a) The adenovirus penton base protein has a cell rounding activity and when penton base was expressed in E. coli, it caused cell rounding and cells adhered to polystyrene wells coated with the protein. Mutant analysis showed that both these properties required an RGD sequence. Virus mutants with amino acid substitutions in the RGD sequence, showed much less adherence to HeLa S3 cells, and also were delayed in virus reproduction (Bai, M., Harfe, B., and Freimuth, P. 1993. Mutations That Alter an RGD Sequence in the Adenovirus Type 2 Penton Base Protein Abolish Its Cell-Rounding Activity and Delay Virus Reproduction in Flat Cells. J. Virol. 67:5198-5205).

b) It has been shown that attachment and entry of coxsackie virus A9 to GMK cells were dependent on an RGD motif in the capsid protein VP1. VP1 has also been shown to bind $\alpha_v\beta_3$ integrin, which is a vitronectin receptor (Roivainen, M., Piirainen, L., Hovi, T., Virtanen, I., Riikonen, T., Heino, J., and Hyypia, T. 1994. Entry of Coxsackievirus A9 into Host Cells: Specific Interactions with a_vb_3 Integrin, the Vitronectin Receptor Virology, 203:357-65).

c) During the course of whooping cough, *Bordetella pertussis* interacts with alveolar macrophages and other leukocytes on the respiratory epithelium. Whole bacteria adheres by means of two proteins, filamentous hemagglutinin (FHA) and pertussis toxin. FHA interacts with two classes of molecules on macrophages, galactose containing glycoconjugates and the integrin CR3. The interaction between CR3 and FHA involves recognition of RGD sequence at the positions 1097-1099 in FHA (Relman, D., Tuomanen, E., Falkow, S., Golenbock, D. T., Saukkonen, K., and Wright, S. D. "Recognitition of a Bacterial Adhesin by an Integrin: Macrophage CR3 Binds Filamentous Hemagglutinin of Bordetella Pertussis." Cell, 61:1375-1382 (1990)).

d) Pertactin, a 69 kDa outer membrane protein of *Bordetella pertussis,* has been shown to promote attachment of Chinese hamster ovary cells (CHO). This attachment is mediated by recognition of RGD sequence in pertactin by integrins on CHO cells and can be inhibited by synthetic RGD containing peptide homologous to the one present in pertactin (Leininger, E., Roberts, M., Kenimer, J. G., Charles, I. G., Fairweather, N., Novotny, P., and Brennan, M. J. 1991. Pertactin, an Arg-Gly-Asp containing *Bordetella pertussis* surface protein that promotes adherence of mammalian cells Proc. Natl. Acad. Sci. USA, 88:345-349).

e) The RGD sequence is highly conserved in the VP1 protein of foot and mouth disease virus (FMDV). Attachment of FMDV to baby hamster kidney cells (BHK) has been shown to be mediated by VP1 protein via the RGD sequence. Antibodies against the RGD sequence of VP1 blocked attachment of virus to BHK cells (Fox, G., Parry, N. R., Barnett, P. V., McGinn, B., Rowland, D. J., and Brown, F. 1989. The Cell Attachment Site on Foot-and-Mouth Disease Virus Includes the Amino Acid Sequence RGD (Arginine-Glycine-Aspartic Acid) J. Gen. Virol., 70:625-637).

It has been demonstrated that bacterial adherence can be based on interaction of a bacterial adhesin RGD sequence with an integrin and that bacterial adhesins can have multiple binding site characteristic of eukaryotic extracellular matrix proteins. RGD recognition is one of the important mechanisms used by microbes to gain entry into eukaryotic cells.

The complete deduced protein sequence of the *Chlamydia pneumoniae* genome was searched for the presence of RGD sequence. There were a total of 54 ORFs that had one or more RGD sequences. Not all RGD containing proteins mediate cell attachment. It has been shown that RGD containing peptides that have proline immediately following the RGD sequence are inactive in cell attachment assays (Pierschbacher & Ruoslahti. 1987. Influence of stereochemistry of the sequence Arg-Gly-Asp-Xaa on binding specificity in cell adhesion. J. Biol. Chem. 262:17294-98). ORFs that had RGD, with proline as the amino acid following the RGD sequence were excluded from the list. Also, RGD sequence may not be available at the surface of the protein or may be present in a context that is not compatible with integrin binding. Since not all RGD- containing proteins are involved in cell attachment, several other criteria were used to refine the list of RGD- containing proteins. A list of ORFs in the *Chlamydia pneumoniae* genome encoding polypeptides with RGD recognition sequence(s) is in the specification.

Non-*Chlamydia trachomatis* ORFs

Chlamydia pneumoniae ORFs were compared to the ORFs in the *Chlamydia trachomatis* genome (French patent applications FR97-15041, filed 28 November 1997 and 97-16034 filed 17 December 1997) using Blastp. Any *Chlamydia pneumoniae* ORF with a Blastp P value worse than e

10 (i.e. $>e^{-10}$) against *Chlamydia trachomatis* ORFs are included in this section. A list of ORFs in the *Chlamydia pneumoniae* genome which are not found in *Chlamydia trachomatis* is set forth above in the specification.

Cell Wall Anchor Surface ORFs

Many surface proteins are anchored to the cell wall of Gram-positive bacteria via the conserved LPXTG motif (Schneewind, O., Fowler, A., and Faull, K.F. 1995. Structure of the Cell Wall Anchor of Surface Proteins in *Staphylococcus aureus*. Science 268:103-106). A search of the *Chlamydia pneumoniae* ORFs was done using the motif LPXTG. A list of ORFs in the *Chlamydia pneumoniae* genome encoding polypeptides anchored to the cell wall is in the specification.

ATCC Deposits

Samples of *Chlamydia pneumoniae* were deposited with the American Type Culture Collection (ATCC), Rockville, Maryland, on November 19, 1998 and assigned the accession number ---. Cells can be grown, harvested and purified, and DNA can be prepared as discussed above. In order to enable recovery of specific fragments of the chromosome, one can run targeted PCR reactions, whose amplification products can then be sequenced and/or cloned into any suitable vector, according to standard procedures known to those skilled in the art.

In addition, a sample of three pools of clones covering chromosomal regions of interest were deposited with the American Type Culture Collection (ATCC), Rockville, Maryland, on November 19, 1998 and assigned the indicated accession number: — . Each pool of clones contains a series of clones. When taken together, the three pools in the sample cover a portion of the chromosome, with a redundancy of slightly more than two. The total number of clones in the sample is 196.

The clones cover the following three regions of interest:

(i) position 30,000 to 40,000 of SEQ ID No. 1, referred to as region A;

(ii) position 501,500 to 557,000 of SEQ ID No. 1, referred to as region B; and

(iii) position 815,000 to 830,000 of SEQ ID No. 1, referred to as region C.

Table 4 lists groups of oligonucleotides to be used to amplify each of ORFs 2-1291 according to standard procedures known to those skilled in the art. Such oligonucleotides are listed as SEQ ID Nos. 1292 to 6451. For each ORF, the following is listed: one forward primer positioned 2,000 bp upstream of the beginning of the ORF; one forward primer positioned 200 bp upstream of the beginning of the ORF; one reverse primer positioned 2,000 bp downstream at the end of ORF, which is 2,000 bp upstream of the end site of the ORF on the complementary strand; and one reverse primer 200 bp downstream at the end of ORF, which is 200 bp upstream of the end site of the ORF on the complementary strand. The corresponding SEQ ID Nos. for the primers are listed in Table 4, where Fp is the proximal forward primer; Fd is the distal forward

primer; Bp is the proximal reverse primer; and Bd is the distal reverse primer. The positions of the 5' ends of each of these primers on the nucleotide sequence of SEQ ID No. 1 are shown in Table 5.

Table 6 lists oligonucleotides (SEQ ID Nos. 6452-6843) to be used to amplify the inserts of each of the 196 clones present in the pooled sample according to standard procedures well known to those of skill in the art. These primers can also be utilized to amplify the chromosomal region corresponding to the region A, B or C within which the particular insert lies. Their positions are indicated in Table 7.

The present invention is not to be limited in scope by the specific embodiments described herein, which are intended as single illustrations of individual aspects of the invention, and functionally equivalent methods and components are within the scope of the invention. Indeed, various modifications of the invention, in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and accompanying drawings. Such modifications are intended to fall within the scope of the appended claims.

All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

WHAT IS CLAIMED IS:

1- An isolated polynucleotide having a nucleotide sequence of a *Chlamydia pneumoniae* genome, comprising

(a) the a nucleotide sequence of SEQ ID No. 1;
(b) the nucleotide sequence contained within the *Chlamydia pneumoniae* genomic DNA in ATCC Deposit No. _______;
(c) the nucleotide sequence contained in a clone insert in ATCC Deposit No. ________;
(d) a nucleotide sequence exhibiting at least 99.9% identity with the sequence of SEQ ID No. 1; or
(e) a nucleotide sequence exhibiting at least 80% homology to SEQ ID No. 1.

2- An isolated polynucleotide which hybridizes to SEQ ID No. 1 or to the *Chlamydia pneumoniae* genomic DNA contained in ATCC deposit No. _______ or to a clone insert in ATCC Deposit No. _______ under conditions of high stringency.

3- An isolated polynucleotide which hybridizes to SEQ ID No. 1 or to the *Chlamydia pneumoniae* genomic DNA contained in ATCC deposit No. _____ under conditions of intermediate stringency.

4- An isolated polynucleotide having a nucleotide sequence of an open reading frame (ORF) of a *Chlamydia pneumoniae* genome, comprising:

(a) a nucleotide sequence chosen from one of ORF2 to ORF 1297;
(b) a nucleotide sequence exhibiting at least 99.9% identity with one of ORF2 to ORF 1297; or
(c) a nucleotide sequence exhibiting at least 80% homology to one of ORF2 to ORF 1297.

5- An isolated polynucleotide which hybridizes to one of ORF2 to ORF 1297 under conditions of high stringency.

6- An isolated polynucleotide which hybridizes to one of ORF2 to ORF 1297 under conditions of intermediate stringency.

7- The polynucleotide of Claims 2, 3, 4, 5, or 6 which encodes the following polypeptides or fragments thereof:

(a) a *Chlamydia pneumoniae* transmembrane polypeptide having between 1 and 3 transmembrane domains;

(b) a *Chlamydia pneumoniae* transmembrane polypeptide having between 4 and 6 transmembrane domains;

(c) a *Chlamydia pneumoniae* transmembrane polypeptide having at least 7 transmembrane domains;

(d) a *Chlamydia pneumoniae* polypeptide involved in intermediate metabolism of sugars and/or cofactors;

(e) a *Chlamydia pneumoniae* polypeptide involved in intermediate metabolism of nucleotides or nucleic acids;

(f) a *Chlamydia pneumoniae* polypeptide involved in metabolism of amino acids or polypeptides;

(g) a *Chlamydia pneumoniae* polypeptide having involved in metabolism of fatty acids;

(h) a *Chlamydia pneumoniae* polypeptide involved in the synthesis of the cell wall;

(i) a *Chlamydia pneumoniae* polypeptide involved in transcription, translation, and/or maturation process;

(j) a *Chlamydia pneumoniae* transport polypeptide;

(k) a *Chlamydia pneumoniae* polypeptide involved in the virulence process;

(l) a *Chlamydia pneumoniae* polypeptide involved in the secretory system and/or which is secreted;

(m) a *Chlamydia pneumoniae* polypeptide of the cellular envelope or outer cellular envelope of *Chlamydia pneumoniae*.

(n) a *Chlamydia pneumoniae* surface exposed polypeptide;

(o) a *Chlamydia pneumoniae* lipoprotein;

(p) a *Chlamydia pneumoniae* polypeptide involved in lipopolysaccharide biosynthesis;

(q) a *Chlamydia pneumoniae* KDO-related polypeptide;

(r) a *Chlamydia pneumoniae* phosphomannomutase-related polypeptide;

(s) a *Chlamydia pneumoniae* lipid A component-related polypeptide;

(t) a *Chlamydia pneumoniae* phosphoglucomutase-related polypeptide;

(u) a *Chlamydia pneumoniae* polypeptide that contains an RGD sequence;

(v) a *Chlamydia pneumoniae* Type III secreted polypeptide;

(w) a *Chlamydia pneumoniae* cell wall anchored surface polypeptide; or

(x) a *Chlamydia pneumoniae* polypeptide that is not found in *Chlamydia trachomatis*.

8- A polynucleotide encoding a fusion protein, comprising one of ORF2 to ORF1297 of Claim 4, 5, or 6 ligated in frame to a polynucleotide encoding a heterologous polypeptide.

9- A recombinant vector that contains the polynucleotide of Claim 1, 2, 3, 4, 5 or 6.

10- A recombinant vector that contains the polynucleotide of Claim 8.

11- A recombinant vector that contains the polynucleotide of Claim 4, 5 or 6, operatively associated with a regulatory sequence that controls gene expression.

12- A recombinant vector that contains the polynucleotide of Claim 8 operatively associated with a regulatory sequence that controls gene expression.

13- A genetically engineered host cell that contains the polynucleotide of Claim 1, 2, 3, 4, 5 or 6.

14- A genetically engineered host cell that contains the polynucleotide of Claim 8.

15- A genetically engineered host cell that contains the polynucleotide of Claim 4, 5 or 6 operatively associated with a regulatory sequence that controls gene expression in the host cell.

16- A genetically engineered host cell that contains the polynucleotide of Claim 8 operatively associated with a regulatory sequence that controls gene expression in the host cell.

17- A method for producing a polypeptide, comprising:

(a) culturing the genetically engineered host cell of Claim 15 under conditions suitable to produce the polypeptide encoded by the polynucleotide; and

(b) recovering the polypeptide from the culture.

18- A method for producing a fusion protein, comprising:

(a) culturing the genetically engineered host cell of Claim 16 under conditions suitable to produce the fusion protein encoded by the polynucleotide; and

(b) recovering the fusion protein from the culture.

19- A polypeptide encoded by the polynucleotide of Claim 4, 5 or 6.

20- The polypeptide of Claim 19 which immunoreacts with seropositive serum of an individual infected with *Chlamydia pneumoniae.*

21- The polypeptide of Claim 19 which comprises the following polypeptides or fragments thereof:

(a) a *Chlamydia pneumoniae* transmembrane polypeptide having between 1 and 3 transmembrane domains;
(b) a *Chlamydia pneumoniae* transmembrane polypeptide having between 4 and 6 transmembrane domains;
(c) a *Chlamydia pneumoniae* transmembrane polypeptide having at least 7 transmembrane domains;
(d) a *Chlamydia pneumoniae* polypeptide involved in intermediate metabolism of sugars and/or cofactors;
(e) a *Chlamydia pneumoniae* polypeptide involved in intermediate metabolism of nucleotides or nucleic acids;
(f) a *Chlamydia pneumoniae* polypeptide involved in metabolism of amino acids or polypeptides;
(g) a *Chlamydia pneumoniae* polypeptide involved in metabolism of fatty acids;
(h) a *Chlamydia pneumoniae* polypeptide involved in the synthesis of the cell wall;
(i) a *Chlamydia pneumoniae* polypeptide involved in transcription, translation, and/or maturation process;
(j) a *Chlamydia pneumoniae* transport polypeptide;
(k) a *Chlamydia pneumoniae* polypeptide involved in the virulence process;
(l) a *Chlamydia pneumoniae* polypeptide involved in the secretory system and/or which is secreted;
(m) a *Chlamydia pneumoniae* polypeptide of the cellular envelope or outer cellular envelope of *Chlamydia pneumoniae*.
(n) a *Chlamydia pneumoniae* surface exposed polypeptide;
(o) a *Chlamydia pneumoniae* lipoprotein;
(p) a *Chlamydia pneumoniae* polypeptide involved in lipopolysaccharide biosynthesis;
(q) a *Chlamydia pneumoniae* KDO-related polypeptide;

(r) a *Chlamydia pneumoniae* phosphomannomutase-related polypeptide;

(s) a *Chlamydia pneumoniae* phosphoglucomutase-related polypeptide;

(t) a *Chlamydia pneumoniae* lipid A component-related polypeptide;

(u) a *Chlamydia pneumoniae* polypeptide that contains an RGD sequence;

(v) a *Chlamydia pneumoniae* Type III secreted polypeptide;

(w) a *Chlamydia pneumoniae* cell wall anchored surface polypeptide; or

(x) a *Chlamydia pneumoniae* polypeptide that is not found in *Chlamydia trachomatis*.

22- A fusion protein encoded by the polynucleotide of Claim 8.

23- The fusion protein of Claim 22 which immunoreacts with seropositive serum of an individual infected with *Chlamydia pneumoniae*.

24- An antibody that immunospecifically binds to the polypeptide of Claim 19.

25- An antibody that immunospecifically binds to the fusion protein of Claim 22.

26- A method for the detection and/or identification of *Chlamydia pneumoniae* in a biological sample, comprising:

(a) contacting the sample with a polynucleotide primer of Claim 1, 2, 3, 4, 5, or 6 in the presence of a polymerase enzyme and nucleotides under conditions which permit primer extension; and

(b) detecting the presence of primer extension products in the sample in which the detection of primer extension products indicates the presence of *Chlamydia pneumoniae* in the sample.

27- A method for the detection and/or identification of *Chlamydia pneumoniae* in a biological sample, comprising:

(a) contacting the sample with a polynucleotide probe of Claim 1, 2, 3, 4, 5, or 6 under conditions which permit hybridization of complementary base pairs; and

(b) detecting the presence of hybridization complexes in the sample in which the detection of hybridization complexes indicates the presence of *Chlamydia pneumoniae* in the sample.

28- A method for the detection and/or identification of *Chlamydia pneumoniae* in a biological sample, comprising:

(a) contacting the sample with the antibody of Claim 24 under conditions suitable for the formation of immune complexes; and

(b) detecting the presence of immune complexes in the sample, in which the detection of immune complexes indicates the presence of *Chlamydia pneumoniae* in the sample.

29- A method for the detection and/or identification of antibodies to *Chlamydia pneumoniae* in a biological sample, comprising:

(a) contacting the sample with a polypeptide of Claim 19 under conditions suitable for the formation of immune complexes; and

(b) detecting the presence of immune complexes in the sample, in which the detection of immune complexes indicates the presence of *Chlamydia pneumoniae* in the sample.

30- A DNA chip containing an array of polynucleotides comprising at least one of the polynucleotides of Claim 1, 2, 3, 4, 5, or 6.

31- A protein chip containing an array of polypeptides comprising at least one of the polypeptides of Claim 19.

32- An immunogenic composition comprising the polypeptide of Claim 19 and a pharmaceutically acceptable carrier.

33- An immunogeneic composition comprising the polypeptide of Claim 20 and a pharmaceutically acceptable carrier.

34- An immunogenic composition comprising the fusion protein of Claim 22 and a pharmaceutically acceptable carrier.

35- An immunogenic composition comprising the fusion protein of Claim 23 and a pharmaceutically acceptable carrier.

36- A pharmaceutical composition comprising the polypeptide of Claim 19 and a pharmaceutically acceptable carrier.

37- A pharmaceutical composition comprising the polypeptide of Claim 20 and a pharmaceutically acceptable carrier.

38- A pharmaceutical composition comprising the polypeptide of Claim 22 and a pharmaceutically acceptable carrier.

39- A pharmaceutical composition comprising the polypeptide of Claim 23 and a pharmaceutically acceptable carrier.

40- A method of immunizing against *Chlamydia pneumoniae*, comprising: administering to a host an immunizing amount of the immunogenic composition of Claim 32.

41- A method of immunizing against *Chlamydia pneumoniae*, comprising: administering to a host an immunizing amount of the immunogenic composition of Claim 33.

42- A method of immunizing against *Chlamydia pneumoniae*, comprising administering to a host an immunizing amount of the immunogenic composition of Claim 34.

43- A method of immunizing against *Chlamydia pneumoniae*, comprising: administering to a host an immunizing amount of the immunogenic composition of Claim 35.

44- A DNA immunogenic composition comprising the expression vector of Claim 11.

45- The DNA composition of Claim 44, wherein the DNA composition directs the expression of a neutralizing epitope of *Chlamydia pneumoniae*.

46- A DNA immunogenic composition comprising the expression vector of Claim 12.

47- The DNA composition of Claim 46, wherein the DNA composition directs the expression of a neutralizing epitope of *Chlamydia pneumoniae*.

48- A screening assay, comprising:

(a) contacting a test compound with an isolated polynucleotide of Claim 1, 2, 3, 4, 5 or 6; and

(b) detecting whether binding occurs.

49- A screening assay, comprising:

(a) contacting a test compound with the polypeptide of Claim19; and

(b) detecting whether binding occurs.

50- A screening assay, comprising:

(a) contacting a test compound with the polypeptide of Claim 22; and

(b) detecting whether binding occurs.

51- A kit comprising a container containing an isolated polynucleotide of Claim 1, 2, 3, 4, 5 or 6.

52- The kit of Claim 51 wherein the polynucleotide is a primer or a probe.

53- The kit of Claim 51 wherein the polynucleotide is a primer and the kit further comprises a container containing a polymerase.

54- The kit of Claim 51 which further comprises a container containing deoxynucleotide triphosphates.

55- A kit comprising a container containing an antibody that immunospecifically binds to the polypeptide of Claim 19.

56- A kit comprising a container containing an antibody that immunospecifically binds to the fusion protein of Claim 22.

Figure 1.

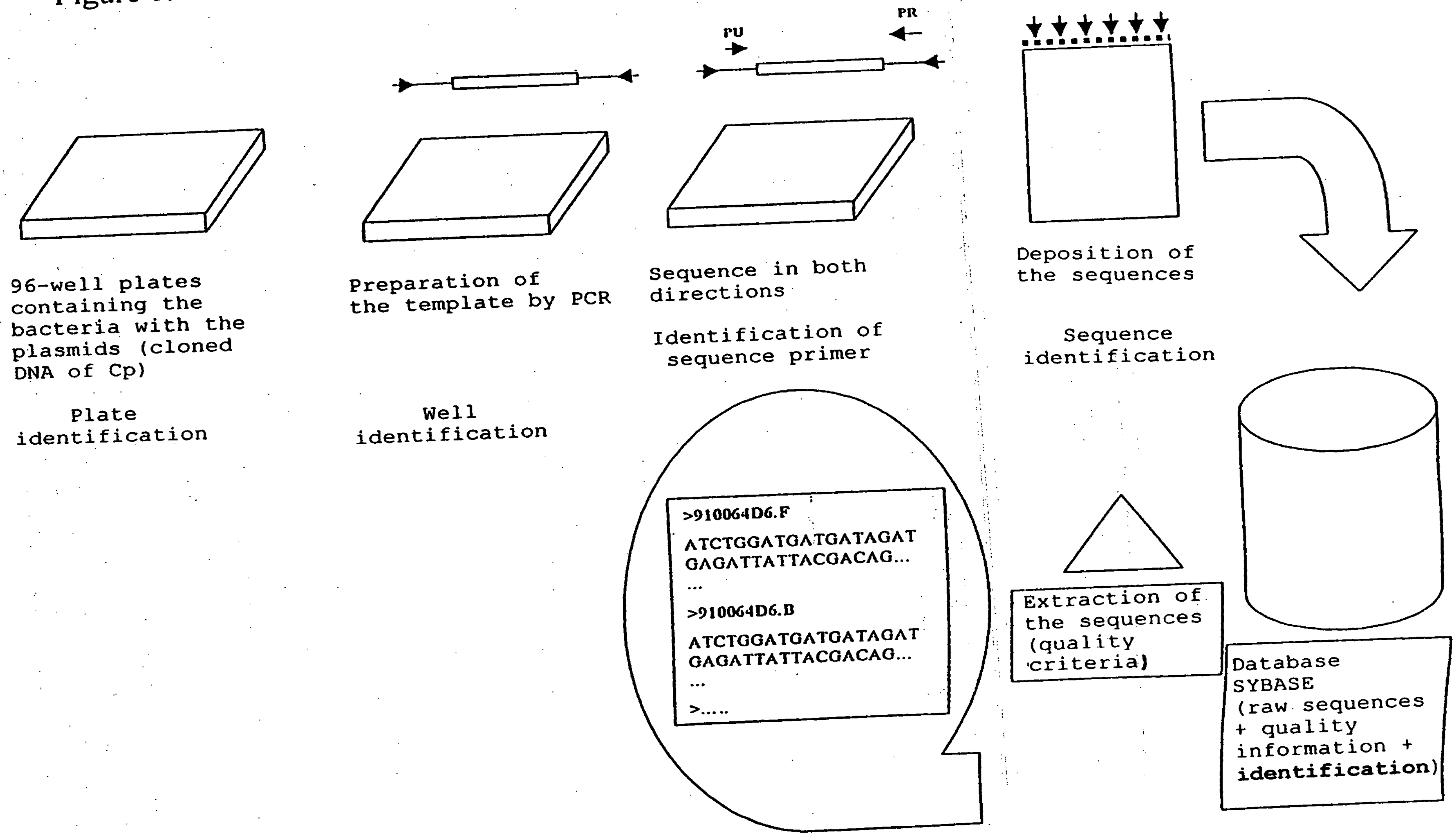

PU
PR
96-well plates containing the bacteria with the plasmids (cloned DNA of Cp)
Plate identification
Preparation of the template by PCR
Well identification
Sequence in both directions
Identification of sequence primer
>910064D6.F
ATCTGGATGATGATAGAT
GAGATTATTACGACAG...
...
>910064D6.B
ATCTGGATGATGATAGAT
GAGATTATTACGACAG...
...
>.....
Deposition of the sequences
Sequence identification
Extraction of the sequences (quality criteria)
Database SYBASE (raw sequences + quality information + **identification**)

Figure 2.

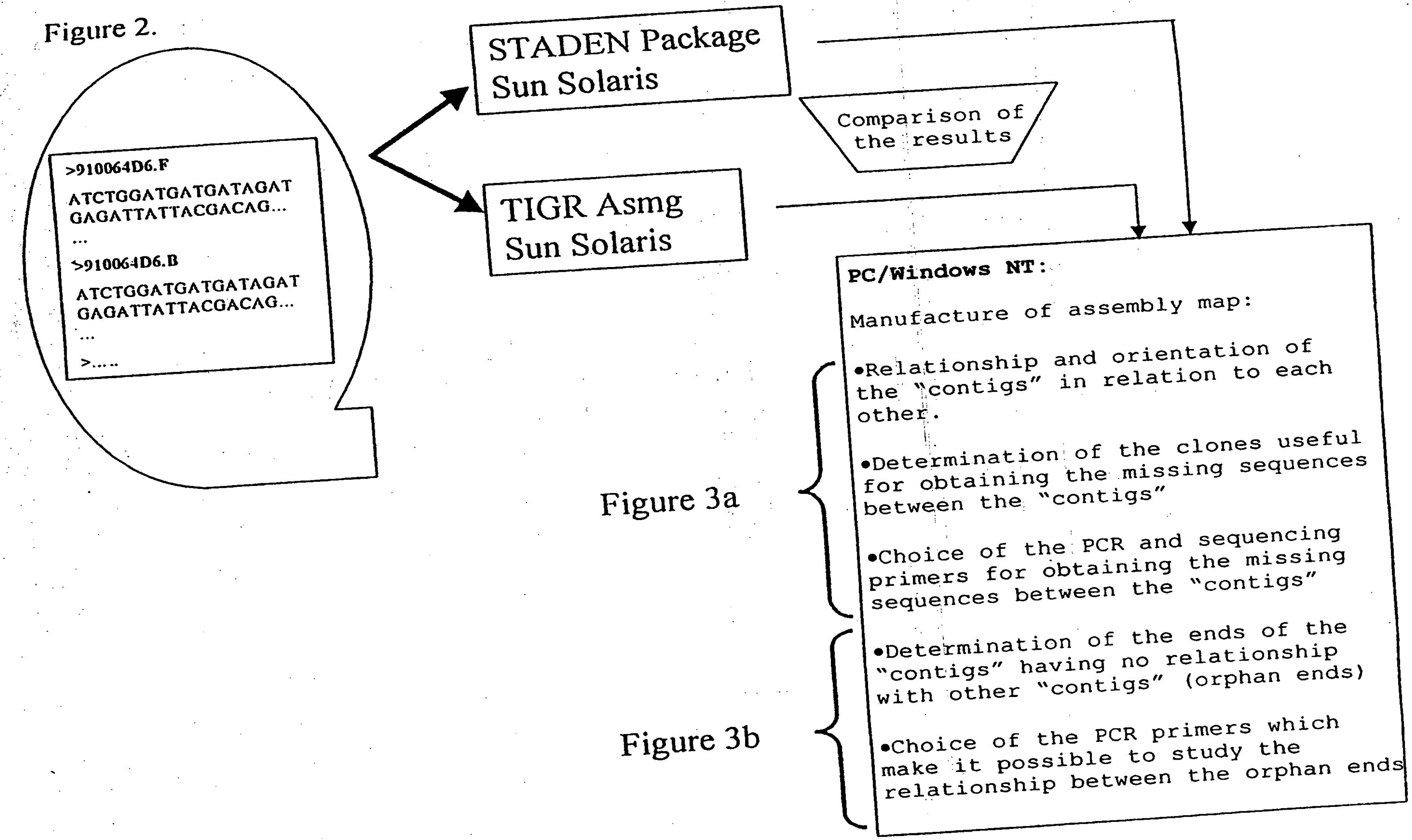
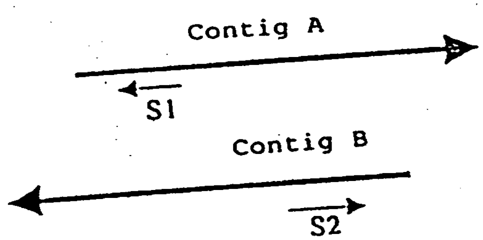


Figure 3a

Figure 3b

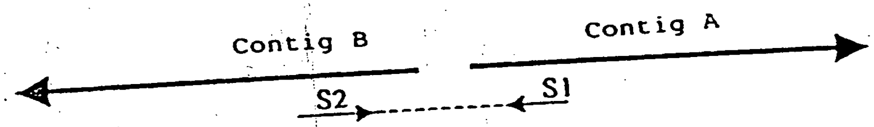
FIGURE 3A



Statistical determination of the sequences:

1- Belonging to the same clone

2- Situated on two different contigs

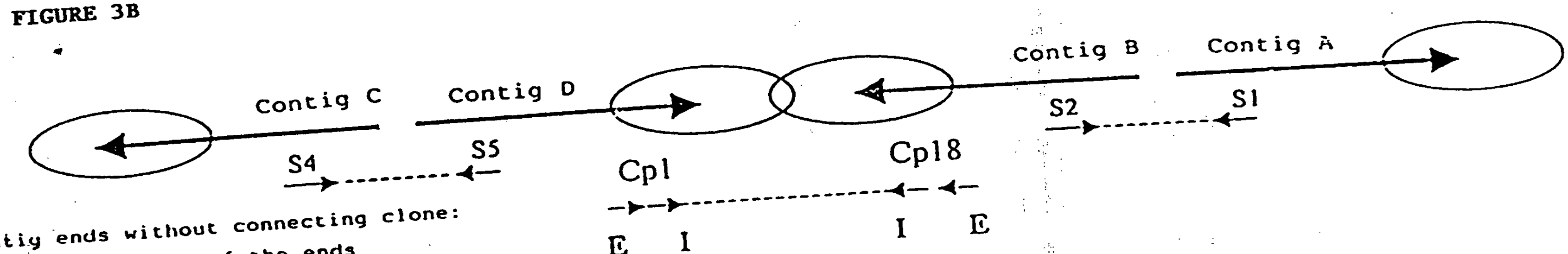


Distribution and position of the sequences on the contigs:

1- Define the arrangement of the contigs in relation to each other

2- Define the PCR primers which make it possible to fill the sequence

FIGURE 3B



Contig ends without connecting clone:

1- Identification of the ends

2- Determination of outer and inner PCR primers for studying the relationships between the contigs

E: outer primers

I: inner primers